Hamid, N. H. and Murray, A. F. and Laurenson, D. and Roy, S. and Cheng, B. (2005) Probabilistic computing with future deep sub-micrometer devices: a modelling approach. In, *IEEE International Symposium on Circuits and Systems, 23-26 May*, pages pp. 2510-2513, Kobe, Japan.

http://eprints.gla.ac.uk/2993/

Probabilistic Computing with Future Deep Sub-Micrometer Devices : A Modelling Approach

Nor H. Hamid*, Alan F. Murray, David Laurenson
Department of Electronics and Electrical Engineering
The University of Edinburgh, Scotland – UK
*nhh@ee.ed.ac.uk

Scott Roy, Binjie Cheng
Device Modeling Group
Department of Electronics and Electrical Engineering
The University of Glasgow, Scotland-UK

Abstract— **An approach is described that investigates the potential of probabilistic "neural" architectures for computation with Deep Sub-Micrometer (DSM) MOSFETs. Initially, noisy MOSFET models are based upon those for a 0.35μm MOS technology with an exaggerated *1/f* characteristic. We explore the manifestation of the *1/f* characteristic at the output of 2-quadrant multiplier when the key n-channel MOSFETs are replaced by "noisy" MOSFETs. The stochastic behavior of this noisy multiplier has been mapped on to a software (Matlab) model of a Continuous Restricted Boltzmann Machine (CRBM) – an analogue-input stochastic computing structure. Simulation of this DSM CRBM implementation shows little degradation from that of a "perfect" CRBM. This paper thus introduces a methodology for a form of "technology-downstreaming" and highlights the potential of probabilistic architectures for DSM computation.**

1. Introduction

Intrinsic MOSFET noise is an increasing problem with no obvious, single solution. As MOSFET dimensions continue to shrink, the effect of noise on the device's reliability and performance becomes more pronounced [1]. The most obvious approaches to a solution are either, to adopt a noise tolerant design approach or more expensively, to try to alleviate the problem through enhanced processing. A more radical approach, hinted at in the technology "roadmap" [2] is to develop architectures that actually exploit the vagaries of Deep Sub-Micrometer (DSM) devices to perform useful computation.

In this paper, we present a systematic simulation study of a probabilistic neural computation architecture using "noisy" MOSFETs. We explore the exploitation of intrinsic MOSFET noise to create useful "stochastic neural" behavior in the context of an intrinsically-interesting probabilistic model developed specifically for hardware-amenability, the Continuous Restricted Boltzmann Machine (CRBM) [3]. The CRBM is based on Hinton's Product of Expert in Restricted Boltzmann Machine form (PoE/RBM) [4], adapted to draw upon the training approach of the diffusion network [5]. We adopt a hierarchical modeling strategy,

whereby we first (Section 3.1) build SPICE-based behavioral models of noisy DSM MOSFETs based upon "atomistic" simulation of the underlying noise mechanisms [6]. We then design the key component of the CRBM, an analog multiplier, using noisy-MOSFET models (Section 3.2). Finally, we embed the noisy behavior of the "noisy-multiplier" in a behavioral model of the CRBM, an example of a probabilistic neural computation architecture (Section 3.3).

2. Motivation

2.1. MOSFET Technology

As MOSFETs shrink and supply voltages reduce, the effect of noise on the devices performance becomes ever more significant [1]. Low frequency noise has been reported [6] of an amplitude comparable with 20-60% of the drain current. Current fluctuations on this scale become a serious performance and reliability issue. To highlight the seriousness of the issue, the Technology Roadmap [2] has suggested some alternative solutions based on emerging research technologies, many of which require either major changes in design discipline or expensive changes in process technology. An alternative suggestion [2] is to use noisy, DSM MOSFETs in an architecture that actually requires them. It is therefore attractive to study computing structures that exploit the unreliable (noisy) behavior of future MOSFETs directly, to implement reliable systems.

2.2. Probabilistic Systems

One possible approach is to use the probabilistic behavior of the MOSFET explicitly, to determine the probability distribution of a circuit output, rather than trying to extract a "deterministic" output. A candidate architecture has been identified in [3], the CRBM. In this architecture, stochastic behavior is introduced through injection of Gaussian noise from an external source [7]. In a hardware implementation, this requires additional circuitry [8] and thus more silicon area. This study asks whether noisy, DSM MOSFETs alone can generate the required stochastic behavior.

3. METHODOLOGY

In this section, we present a systematic approach that embeds noisy, DSM MOSFET models in a probabilistic neural computation architecture. The CRBM is used here primarily as an interesting example of this the form of architecture. The CRBM was chosen for its simple and hardware-amenable algorithm [8].

3.1. Modelling Noisy MOSFETs

A good, computationally-simple, noisy MOSFET model is critical to this work. The model must represent the predicted behavior of a noisy DSM MOSFET. At this concept-proving stage, however, simplicity is more important than detailed accuracy. It is not, therefore, our objective in this study to develop a complex, arbitrarily-accurate noisy DSM MOSFET model. Rather, we have developed a model that produces the correct *form* of noise behavior, valid in all operating regimes, and easy to implement in circuit simulator.

3.1.1. Noise Model

Probabilistic neural computation is generally aimed at dealing with real-time data. In embedded systems, one area where DSM technology may be highly desirable for reasons of size, data often encodes biological or even audio information, of relatively low bandwidth [2]. We therefore focus for now on low frequency noise models, i.e. flicker noise and Random Telegraph Signal (RTS) noise. In this paper, we focus on flicker noise, for which we have good results. RTS will be discussed in a later paper, when statistically-significant results are available.

Flicker noise, or 1/f noise, is conveniently described by its' drain current Power Spectral Density (PSD) [9]:

$$S_I = \frac{K_f I_{ds}^{a_f}}{C_{ox} L_{eff}^2 f^{e_f}} \quad (1)$$

where K_f, and a_f are process-dependent constant parameters. C_{ox} is the oxide capacitance and L_{eff} is the effective channel length. Flicker noise frequency exponent e_f dependence on gate bias has been reported in [10]. From (1), we can predict the behavior of flicker noise with varying bias conditions, frequencies, and physical dimensions.

3.1.2. Modeling Methodology

The noisy MOSFET was modeled in a conceptually simple manner, by augmenting a conventional MOSFET model with a noise source, $n(t)$ as shown in Fig. 1a. The noise source $n(t)$ thus represents *time domain* flicker noise. This flicker noise was generated from the PSD described by (1) using a *Sum-of-Sinusoids* technique adapted from that in [11]. Some limitations of this technique are discussed in [12], but it is good enough for our purpose – to provide a computationally-tractable noise model that can be used in transient analysis to provide a reasonably accurate representation of noise behavior and response. The development and implementation of a fast, efficient, and highly-accurate noisy MOSFET is beyond the scope of this study.

3.1.3. Noisy MOSFETs: Result and Discussion

The noisy MOSFET model is based upon that for a 0.35μm AMS technology n-channel MOSFET. For 0.35μm technology, the actual noise generated according to (1) is too low in amplitude to be interesting. To enhance the noise source $n(t)$ effect, we simply multiplied the flicker noise co-efficient K_f by 10^5, to emulate the effect of the changes introduced by DSM device shrinkage and to produce a noise level that is significant in this context. Fig. 1b shows the drain current transient characteristic (drain voltage swept from 0 to 3.5 Volts) of a noisy n-channel MOSFET with a fixed gate voltage of 2 Volts. The underlying I-V characteristic of the noisy MOSFET is overlaid upon that of a "noiseless" n-channel MOSFET. We have verified the generated noise PSDs and confirmed that the noise generated has, indeed, true *1/f* characteristics.

3.2. A Noisy, 2-Quadrant Analog Multiplier

Multipliers are used extensively in almost all forms of neural architecture, representing, primarily, the effects of synaptic gating. Potentially, therefore, an intrinsically noisy analogue multiplier could replace the external noise injection in a CRBM neuron.

A 2-quadrant Chible multiplier (Fig. 2) was chosen for its simple architecture [13] and for the wide range of values of V_w (0 to V_{DD}) over which the output remains linearly proportional to the input. For the noisy multiplier implementation in this paper, we replaced MOSFETs M1, M4 and M5 with noisy MOSFETs, to introduce noise to the signal path without, at this stage, compromising the performance of the current mirrors.

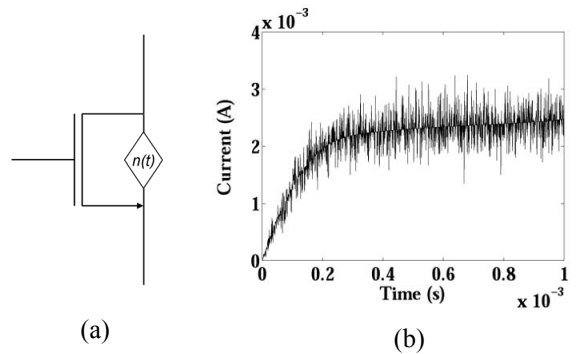Using time-domain transient analysis, we simulated the noisy 2-quadrant multiplier (Fig. 3).

(a)   (b)

Figure 1. (a) Noisy n-channel MOSFET. (b) Noisy drain current for Vgs=2.0V and Vds=0-3.5V simulated for 1ms.
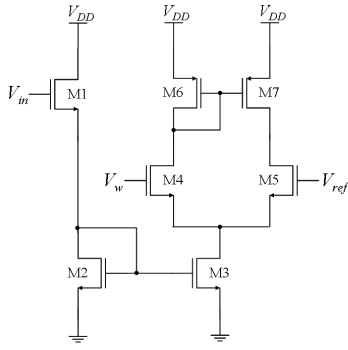
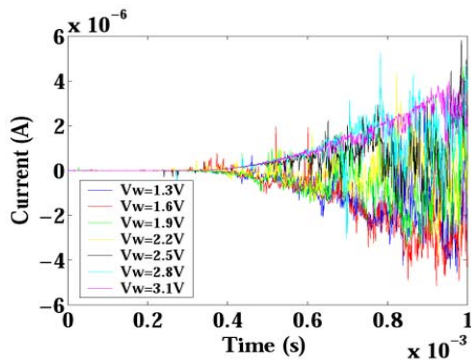Figure 2.   2-Quadrant Chible Multiplier

Figure 3.   Noisy 2-quadrant Chible multiplier implemented by replacing M1, M4, and M5 with noisy n-channel MOSFETs.

The PSD of the output current was analyzed by simulating the noisy multiplier for several combinations of input voltages, V_w (0V, 0.5V, 3.5V) and V_{in} (2.0V, 0.1V, 2.5V). The noise at the output of the multiplier also exhibits a *1/f* characteristic. A PSD was then generated for each bias combination (V_w and V_{in}). The relation between bias combinations and PSDs was then interpolated using the heuristic power-law function:

$$PSD = \frac{a}{f^b} \qquad (2)$$

where *a* and *b* are both fitting parameters. Comparing (2) with (1), "*a*" can be correlated to $K_f I_{ds}^{af}/C_{ox}L_{eff}^2$ and "*b*" the flicker noise frequency exponent e_f.

It can thus be seen that the output of the noisy multiplier exhibits *1/f* characteristic. The fitting parameters of the noisy multiplier PSD were correlated to physical/process parameters and drain current found in (1). The power-law function (2) gave a good generalized relationship between input voltages and PSDs, and is thus usable as an empirical model of the noise characteristics of the same multiplier across its operating range.

3.3. DSM noise and Probabilistic Computation

The principle objective of this study is to explore whether noisy DSM MOSFETs have the potential to provide the stochastic behavior needed by a probabilistic architecture such as the CRBM.

3.3.1. The CRBM : Overview

The CRBM is a probabilistic neural computation model capable of modeling analog data, and adapted ("trained") according to a simple, unsupervised training algorithm [3,7]. It has been shown that the CRBM is amenable to VLSI implementation and potentially useful as both a robust classifier and as a "novelty detector" [7,8]. Reference [3] and [7] provide a full description of CRBM architecture, algorithm, and training procedure.

The basic building block of CRBM is the continuous stochastic neuron [8], shown in Fig. 4. The external Gaussian noise injection provides the stochastic behavior of the neuron [7]. Adapting the detailed injection of noise allows CRBM neurons to adapt during training to become binary-stochastic, deterministic, or continuous-stochastic [3]. This is one of the most interesting features of the CRBM and enhances its modeling ability significantly [3]. In this work, we aim to eliminate the need for external Gaussian noise injection by introducing intrinsic DSM-style fluctuations, initially in the multiplier blocks of CRBM.

3.3.2. A CRBM with injected DSM noise

The stochastic behavior of the neurons was introduced by replacing the normal multiplication process by the noisy multiplication developed in the last section, removing the explicit injection of Gaussian noise in the underlying CRBM model. As the noisy multiplier was implemented in hardware, we must map the input parameter and output characteristics carefully to match that of the software CRBM on which this study draws.

For each combination of inputs (S_i and ω_i), a corresponding PSD is generated using the partly-empirical fitting function (2) and fitting parameters *a* and *b*. Using this PSD, a noise datum was then generated using the sum-of-sinusoid technique [11]. The noise datum was then added to the multiplier output in what was now a software (Matlab) simulation of a hardware CRBM.
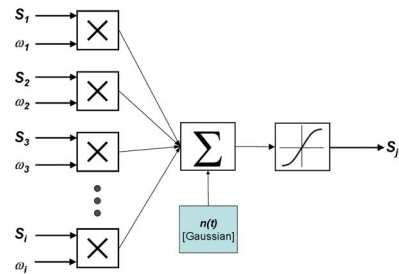

Figure 4.   CRBM neuron with external Gaussian noise *n(t)*.

3.3.3. Results and Discussion

To begin to explore the effects of DSM noise on probabilistic architectures, the original "perfect" software CRBM and the model of a hardware CRBM with DSM noise injected were both trained to model the same two, initially well-separated, clusters of 2-dimensional data shown in Fig. 5a. Then, the trained CRBMs were presented with a random data. After 20-steps of data reconstruction, both CRBMs showed a good reconstruction of the original cluster (see Fig. 5b and Fig. 5c).

This initial, preliminary result shows that stochastic neuron behavior introduced through noisy MOSFET can produce CRBM performance comparable with that of a "perfect" CRBM with explicit noise injection. Future studies will explore more complex, higher-dimensional data to probe the limitations of this DSM-noise-induced modeling ability.

4. CONCLUSION AND FUTURE WORK

We have presented a principled, but pragmatic approach that including the effect of noise in MOSFETs that do not yet exist into circuit models for probabilistic architectures. The approach is based upon detailed atomistic (physical) models of DSM devices, abstracted into a form that is suitable for inclusion in both circuit-level and architecture-level simulation. Although we have chosen to study a probabilistic architecture, the approach is perfectly generic and could be used to explore the effects of DSM noise on conventional computational paradigms, significantly in advance of the availability of ICs with true DSM devices. We have also presented the first study of a probabilistic architecture that actually uses the intrinsic noise in DSM devices directly as part of its probabilistic functionality. Early results suggest that, while noise levels in current commercially-available MOS technology neither present a serious problem nor offer a route to efficient probabilistic computation, those in DSM devices will do both. We do not claim to have proved that DSM MOSFET noise can be exploited in all contexts and with all data, for probabilistic computation. We have, however, shown for the first time that DSM MOSFET noise has the potential to be exploited for probabilistic neural computation architecture hardware implementation.

When MOSFET dimensions fall beyond 50nm [6], Random Telegraph Signal (RTS) noise becomes dominant at low frequency. In ongoing future work, we model noisy MOSFETs with intrinsic RTS noise and implement them in CRBM using the same approach reported in this paper.
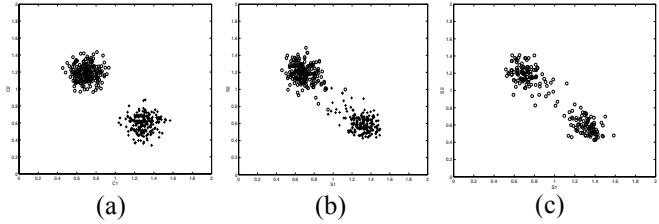
Figure 5. (a) Training data cluster for both CRBMs. (b) 20 step reconstruction of CRBM injected external gaussian noise. (c) 20-steps reconstruction of CRBM implemented with noisy multiplier model.

REFERENCES

[1] G. Ghibaudo, and T. Boutchacha, "Electrical noise and RTS fluctuation in advanced CMOS devices," *Microelectronics Reliability*, vol. 42, 2002, pp. 573-582.

[2] Semicondutor Industry Association (SIA), *International Technology Roadmap for Semiconductor 2003 Edition.* Available: http://public.itrs.net.

[3] H Chen, and A. F. Murray, "A Continuous Restricted Boltzmann Machine with a hardware-amenable learning algorithm," in *Proc. of the International Conference on Artifical Neural Networks (ICANN2002)*, 2002, pp. 358-363.

[4] G. E. Hinton, "Training Product of Expert by minimizing contrastive divergence," *Neural Computation*, vol. 14, no.8, 2002, pp. 1771-1800.

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

[6] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuation in Decananometer and Nanometer-scale MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 9, Sept 2003, pp. 1837-1852.

[7] H. Chen and A. F. Murray, "A Continuous Restricted Botlzmann Machine with an implementable training algorithm," in *Proc. IEE. Vision, Image, and Signal Processing*, vol. 150, 2003, pp 153-158.

[8] H. Chen, P. Fleury, and A. F. Murray, "Minimizing Contrastive Divergence in noisy mixed-mode VLSI neurons," in *Advances in Neural Information Processing System (NIPS2003)*, vol. 16, 2004, In Press.

[9] BSIM3v3.2.2 Mosfet Model: User' Manual. Available: http://www-device.eecs.berkeley.edu/~bsim3.

[10] W. Y. Ho, C. Surya, K. Y. Tong, W. Kim, A. E. Botcharev, and H. Morkoc, "Characterization of flicker noise in GaN-based MODFET's at low drain bias," *IEEE Trans. Electron Devices*, vol. 46, June 1999, pp. 1099-1104.

[11] P. Bolcato and R. Poujois, "A new approach for noise simulation in transient analysis," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1992, pp. 887-890.

[12] A. Demir, E. W. Y. Liu, and A. L. S-. Vincentelli, "Time-domain non-Monte Carlo noise simulation for non-linear dynamic circuits with arbitrary excitations," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, May 1996, pp. 493-505.

[13] H. Chible, "Analog circuits for synapse neural networks VLSI implementation," in *The 7th IEEE Int. Conf. on Electronics, Circuits, and Systems (ICECS 2000)*, vol. 2, 2000, pp. 1004-1007.

[14] H. Chen, "Continuous-valued Probabilistic Neural Computation in VLSI," Ph.D. Thesis, School of Eng. and Elect., University of Edinburgh, Edinburgh, UK, 2004.