

MicroBayes : Probabilistic Machines for Low-level Sensor Interpretation

Projet d'équipe-action déposé auprès de l'action PCS
(Pervasive Computing Systems) du Labex Persyval

March 31, 2017

Contents

1	Members	2
2	Principle and objectives	2
3	Scientific challenges and contributions	3
4	Research directions and methodology	5
5	Expected results	9
6	Calendar, milestones and Gantt chart	10
7	Requested funds	11
8	Annex A: Explicit response to the initial committee	12
9	Annex B: PhD Thesis proposal	12
10	Annex C: Post-Doc proposal	13

1 Members

Emmanuel Mazer (DR CNRS **LIG**) emmanuel.mazer@imag.fr

Laurent Girin (Prof. Grenoble-INP **GIPSA-lab**) laurent.girin@inria.fr

Laurent Fesquet (MCF Grenoble-INP **TIMA**) laurent.fesquet@imag.fr

Didier Piau (Prof. UJF **Institut Fourier**) didier.piau@univ-grenoble-alpes.fr

2 Principle and objectives

The development of modern computers is mainly based on increase of performances and decrease of size and energy consumption. This incremental evolution is notable, but it involves no notable modification of the basic principles of computation. In particular, all the components perform deterministic and exact operations on sets of binary signals. These constraints obviously impede further sizable progresses in terms of speed, miniaturization and power consumption. As detailed below, the goal of the MICROBAYES project is twofold:

- to investigate a radically different approach to perform computations, namely stochastic computing using stochastic bit streams.
- to show that stochastic architectures can outperform standard computers to solve complex inference problems both in terms of execution speed and of power consumption. We will demonstrate the interest and feasibility of stochastic computing on two applications involving low-level information processing from sensor signals, namely sound source localization and separation.

This radical change of current computation models, at their deepest design level, may very well lead to the elaboration of low power reactive systems directly connected to their environment through noisy sensors. As such, we think the approach of this project fits in the theme **PCS** of the **Labex Persyval-lab**.

The use of non-standard architectures to process uncertain information is currently experiencing a remarkable growth, see for example the MIT Probabilistic Computing Project¹ and the DARPA projects PPAML² and UPSIDE³. A European counterpart is the European project BAMBI⁴ (Bottom-up Approaches to Machines dedicated to Bayesian Inference), which includes the LIG laboratory and ends in 2016. The BAMBI project revealed the huge potential of stochastic computing when considered at the bit level. This potential is even further reinforced by advances in massive generation of random bits by dedicated hardwares

¹<http://probcomp.csail.mit.edu/>

²<http://www.darpa.mil/program/probabilistic-programming-for-advancing-machine-learning>

³<http://www.darpa.mil/program/unconventional-processing-of-signals-for-intelligent-data-exploitation>

⁴<https://www.bambi-fet.eu/fr/>

(magnetic tunnel junctions [?], entropy extractors [? ? ?] or standard image sensing devices [?]), which open the way to the design of truly stochastic computers in the very short term.

In the present MICROBAYES project, our aim is to develop a stochastic computing architecture adapted to the resolution of a family of Bayesian signal processing problems. In such problems, the goal is to estimate the posterior probability distribution of an unknown targeted information given a set of measured signals provided by low-level sensors. For example, one might want to find the probability distribution of the position of a sound source given some acoustic signals captured by a set of microphones. As is well known, the analytical expression for this posterior distribution is, in general, intractable and one alternative is to output samples drawn out of this distribution using sampling techniques such as the Gibbs sampler. Then, one can extract the moments of interest such as the mean and the variance (in this example, an average estimate of the sound source position and the precision of this estimation).

The MICROBAYES project is concerned by the design of machines producing such samples. The envisioned architecture is based on Markov Chain Monte Carlo (MCMC) methods. It is made of two modules: the *proposal* module and the *selection* module (see Figure 1). To produce a new sample the proposal module devises a possible value (C) for the next sample. The selection module then makes a binary choice: either to output the new proposed sample or to duplicate and output the previous sample. This binary decision is based on a random draw on a Bernoulli distribution of parameter p . Of course the parameter p and the proposed new sample must be carefully chosen to produce samples that are effectively drawn from the targeted posterior distribution. One key feature of MICROBAYES is to perform this random draw with a dedicated hardware which is both fast and energy efficient.

3 Scientific challenges and contributions

Our project targets three main scientific and technical contributions, each being a challenge:

- (a) To design non von Neumann architectures dedicated to the processing of stochastic bit streams.
- (b) To study new algorithms for stochastic inference based on generating sets.
- (c) To evaluate stochastic machines on difficult Bayesian inference problems related to low-level sensor fusion.

The first contribution of MICROBAYES is to implement a stochastic machine to decide between the new proposed sample or the previous sample. To do so, we will combine several stochastic bit streams into a pipeline of logical gates to obtain one single stochastic bit stream

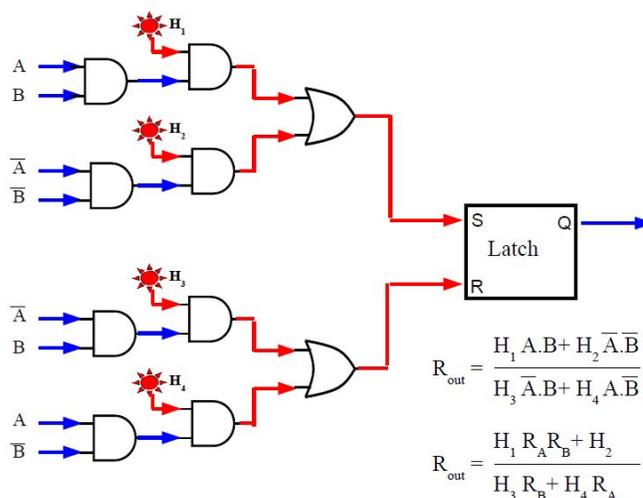


Figure 1: The sampling register is made of a quantized version of the variables of the posterior distribution, for example, 3D-coordinates (X, Y, Z) of the source in the localization problem. The inputs of the machine are pre-processed quantized values of the sensor readings, for instance inter-channel time / phase / level differences (ITD, IPD, ILD). These inputs, the current value of the sampling register (X, Y, Z) , and the proposed new sample C are used to address the parameters needed to generate the stochastic bit streams powering the stochastic machine. The stochastic machine is a cascade of logical operators (such as Muller C elements) producing a single stochastic binary signal, whose first output is used to decide to select or reject the proposed sample. The proposal module can be parametrized by setting the generating sets.

as output. The relative proportion of *ones* in this bit stream is an unbiased estimator of the parameter p (a fundamental property of stochastic computing). The first bit coming out of this pipeline is used to make the decision: For a *one*, we use the new proposed sample as the new sample, for a *zero*, we output the same sample as in the previous step. The main advantages of these stochastic machines over standard processors (such as GPU) are their speed (one clock cycle is often enough to decide) and their bitwise design leading to very low power consumption architectures.

The second contribution of MICROBAYES is to parameterize the proposal module by some *generating sets*. A generating set is a finite collection of admissible displacements, which characterize the moves the algorithm may make at each step to generate a new element of the sampling space. For example, given some quantization on n bits for the variables of the posterior distribution, one may consider generating sets of the cyclic group $\mathbb{Z}/2^n\mathbb{Z}$. Using a unique generating set or a combination of generating sets to propose new samples still ensures the convergence of the algorithm. For example two generating sets may be interlaced to explore the sampling space and to adapt the sampling strategy to a particular

problem.

The third contribution of MICROBAYES is to demonstrate that the proposed architectures can be applied to solve inference problems in real-world low-level signal processing situations. In particular, a challenge in this context is to design architectures that are robust to a large number of inputs. Indeed, stochastic machines deal with probability densities of discrete variables, whereas the problems we address are naturally expressed by continuous variables. The quantization of the variables is thus a major issue. We plan to address this issue in two steps of increasing complexity. First we shall consider the sound source localization problem for which the input and output variables may be quantized with a relatively low resolution. Then we will consider the source separation problem, i.e. recovering individual source signals from a mixture of them, where the quantization issue is more problematic.

The second and third challenges above are linked: MICROBAYES will have to demonstrate that generating sets are good candidates to parameterize a generic stochastic machine as well as to provide efficient sampling strategies for the more specific source localization and separation problems.

4 Research directions and methodology

The project is organized into four modules:

- M1 Architecture and compilers for stochastic machines (LIG, TIMA)
- M2 Algorithms for stochastic search (Institut Fourier, LIG)
- M3 Stochastic algorithms for the localization and separation of sound sources (GIPSA-lab, LIG, Institut Fourier)
- M4 Evaluation on simulated and real-world acoustic signals (GIPSA-lab, LIG, TIMA)

These modules are complementary and interrelated. For example, M2 and M3 will collaborate to find strategies to explore the posterior distributions in the case of sound processing. M1 will define compilers for the models defined in M3. These models will be implemented and tested in M4, in simulation and on real machines.

4.1 M1 Architecture and compilers for stochastic machines (LIG, TIMA)

The goals of M1 are:

1. To produce a hardware description program in VHDL (VHSIC Hardware Description Language, where VHSIC stands for Very High Speed Integrated Circuit) of the architecture described in Figure 1, to simulate this design in MODELSIM and finally to synthesize a working probabilistic machine on an FPGA (Field Programmable Gate Array).
2. To produce compilers taking as input Bayesian Programs and targeting this machine. The purpose of the compiler is to fill the memory of the machine (see Figure 1) with the appropriate probability values p . When running the machine these values will be used to synthesize the stochastic bit streams set as input to the logical pipeline.

This design will be inspired by the architecture proposed by Marvin Faix for his ongoing PhD [?]: A generic Bitwise Gibbs Sampler. This architecture will be modified

1. To allow the configuration of this compiler with search strategies given as sequence of generating sets (in connection with M2).
2. To improve the logical elements of the machine to better estimate the multiplication between odds⁵ with stochastic arithmetics by adding memories to the Muller C element.
3. To include massive source of entropy (random bit generators) to increase the speed of the machine by using entropy multipliers, TRNG (true random generator) connected to bank of LFSR (Linear Feedback Shift Register).

Given an inference problem, the Bayesian Programming formalism allows the programmer to specify variables, conditional probability distributions, joint distributions and the desired inference (i.e. the posterior distribution). PROBT, a Bayesian programming language developed at LIG, will be used in M3 to specify the source localization and separation inferences. The compiler will transform these programs to fill the memory of the probabilistic machine with the parameters necessary to generate the stochastic bit streams coming into play when solving the desired inference.

4.2 M2 Algorithms for stochastic search (Institut Fourier, LIG)

During the BAMBI project, we gained experience in inference algorithms for bit wise stochastic machines. Contrary to other methods, which often boil down to the use of fixed or floating point arithmetic on an array of processors [? ?], in this approach one studies in details the use of simple bitwise operations to make exact and approximate inference. This led to several interesting and efficient ways to perform inference. For example, the BM1 machine (Figure 4.2) is based on the exploitation of stochastic buses: by counting the number of

⁵For a binomial distribution p the odd is defined as $\frac{p}{1-p}$

zeroes and ones on each line of the bus, one encodes a discrete probability distribution on a set whose size is the dimension of the bus. One can then combine these buses with simple AND gates, to obtain very efficient Bayesian filters which could already outperform standard computers on simple but useful problems, for example pseudo-noise acquisition sequences used to synchronize antennas in telecommunications.

However such an approach does not a priori resist the curse of dimensionality phenomenon. To avoid this, a BM2 machine was designed, based on bitwise Gibbs sampling. This machine was shown to be able to solve some otherwise intractable problems. However, as is well known in Gibbs sampling, this type of algorithm is deemed to fail when some flat probabilistic landscapes must be crossed before one reaches regions with high enough probability mass. This is where the idea of *generating sets* comes into play, as a way to explore efficiently the search space.

A version of this idea called RRT (for *rapidly exploring random trees*) [?], which could prove useful in our context, is to alternate between large exploration steps, designed to jump quickly to other regions of the landscape, and some steps of standard size, designed to examine more closely each of these regions. In these so-called tree expansion strategies [?], the algorithm grows a tree rooted at the start configuration aiming at the goal configuration, by repetitively expanding it in random directions. The efficiency of these methods have been demonstrated on some large classes of motion planning problems, and the failure probability and its dependence on some characteristics of the landscape have been rigorously characterized in some cases [?]. However, much work remains to be done to advance their theoretical understanding.

To sum up, our goal, by exploring families of generating sets and selecting appropriate ones, will be to devise efficient versions of the Gibbs algorithm suited to our setting. Devising within M3 generating sets that are appropriate for the sound source localization and separation problems is of major interest to us. These selected generating sets will then be used to parameterize the compilers produced in M1.

4.3 M3 Stochastic algorithms for the localization and the separation of sound sources (GIPSA-lab, LIG, Institut Fourier)

We propose two possible applications for the technology developed in this project, in the field of sound processing for natural scene analysis using multi-microphone recordings: sound source localization (SSL) and sound source separation (S3) (i.e. recovering the unobserved individual source signals from the recorded mixture of them). Within the last decade, both tasks have been extensively addressed within a probabilistic framework [? ? ? ? ?]. This is because both sound source signals and their propagation process are subject to many

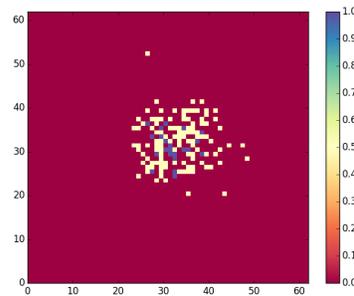
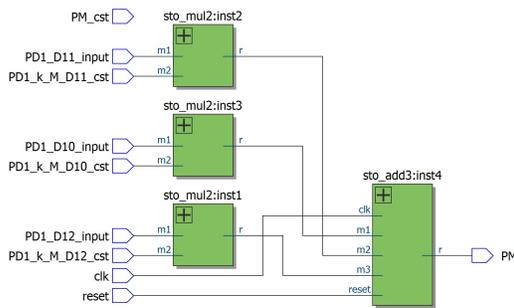
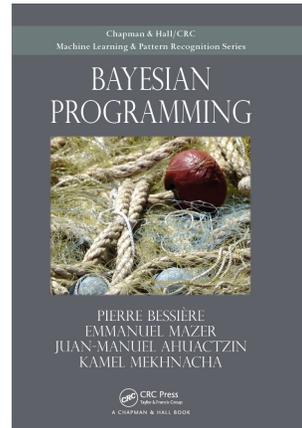
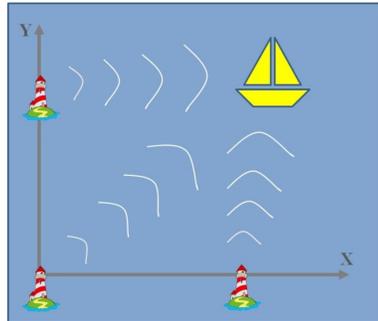


Figure 2: BM1 [? ?] synthesis and test. A localization problem with sensor fusion (top left) is translated into an inference problem using a Bayesian programming language [?] (top right). A compiler automatically generates the stochastic circuit (bottom left) performing the inference. In this simple example, the mean position of the boat (bottom right), obtained after only one clock cycle, is within 2% of the actual location. The MICROBAYES project plans to extend this “toy” localization problem to the realistic and much more complex case of sound sources in natural environment.

fluctuations and uncertainties. SSL methods rely on a probabilistic model of the distribution of acoustic features (inter-channel time/phase/level differences: ITD/IPD/ILD) as a function of source location. S3 methods rely on a probabilistic model of the distribution of the mixture signal as a function of elementary source signals. In both cases, the information of interest is inferred from the measured signals using the corresponding posterior distribution model. In most cases, *this posterior distribution has no closed-form expression*. Moreover, especially in S3 problems, the amount of information to be inferred is huge: Several matrices of source time-frequency coefficients, with frequency dimension that is hundreds of bins and time dimension that is about a hundred frames per second of signal [? ?]. Sampling techniques were proposed for S3 in the literature [? ? ?], but did not meet success. They remain limited to simplified “laboratory” cases, due to the prohibitive amount of calculations involved in realistic scenarios. Approximate closed-form solutions are preferred, using approximation

techniques, e.g. variational methods [?]. The proposed project can enable to avoid the limitations of approximation methods. It can lead to a revival for the sampling techniques in a S3 task, giving the opportunity to exploit the power of these methods in realistic use cases, something that was simply impossible so far.

The Bayesian programming language PROBT will be used to express the SSL and S3 algorithms. Then the compilers produced in M1 will be parameterized by selecting appropriate sequences of generating sets to efficiently explore the sampling space of SSL and S3. These sequences will be developed in collaboration with M2. As for the SSL and S3 algorithms themselves, we will start from existing algorithms and adapt them or develop new ones, so that to optimally fit the requirements and characteristics of the inference process from M2 and implementation on the machine from M1.

4.4 M4 Evaluation on simulated and real-world acoustic signals (TIMA, GIPSA-lab)

The goal of this module is to test and evaluate the implementation of stochastic machines on FPGA with actual signals obtained from both acoustic simulations and a real-world experimental acoustic test bed. As our primary goal is to evaluate the stochastic inference engine, we will allow standard computers to perform some pre- and post-processing. For example, acoustic feature extraction (ITD/IPD/ILD for SSL and time-frequency coefficients for S3) will be computed offline.

Several designs will be tested and compared to each other and with more standard methods. We will evaluate the inference performances (in terms of localization and signal reconstruction accuracy) for the different strategies designed in M1, M2 and M3, e.g. different quantization resolutions of the continuous variables at stake, and different space exploration strategies. Energy consumption will be estimated and compared with the energy consumption of standard machine performing the same task. The M4 module will also be used to draw the roadmap for future developments of stochastic machines.

5 Expected results

Technically our goal is to demonstrate a working stochastic processor performing sound source localization and separation from a set of (noisy) acoustic signals. A success would open new perspectives in sensor processing, leading to the possibility to use inexpensive and noisy sensors on devices working with extremely low energy resources. Several compilers will be developed to target several stochastic processors and transform Bayesian programs into dedicated bitwise Gibbs samplers. New designs will also be devised to build these machines. From a theoretical standpoint we will characterize the search strategies in high-dimension

spaces by studying several generating sets of $[0, 1]^n$. In other words, we would like to identify the heuristics used to explore large search spaces and knowingly apply them to NP problems in a more abstract and possibly more efficient way.

We are currently working on an industrial setup to exploit stochastic computing with the early results of the BAMBI project. This setup will include several CNRS laboratories (LIG, IEF, Numphi, ISIR), one of its spinoff companies (PROBAYES), and defense (MBDA) and aerospace contractors. Indeed, getting low power signal processors dealing with noisy sensors is of major interest for the defense and aerospace industries since some recent tests made on the BM1 by TIMA demonstrate the robustness of probabilistic processors with respect to bit flips due to electromagnetic radiations. By providing new results and by developing some know-how in this emerging field, the MICROBAYES project could greatly contribute to push forward this new scientific and industrial endeavor.

Naturally, we expect from this project many opportunities to publish the developments and results in the best international conferences and international journals related to the topics, from theoretical statistics to applied audio processing.

6 Calendar, milestones and Gantt chart

The total project duration will be 3 years.

A PhD student will be recruited, funded by the project, and will have three tasks: (a) to understand and implement in hardware algorithms developed in collaboration with D. Piau and E. Mazer (M1, M2); (b) To develop compilers to transform the Bayesian programs written in M3 to target the architecture proposed in M1; (c) To analyze the performances of the proposed stochastic machines in terms of power consumption, speed and surface.

During year one, the action team will start by hiring the PhD student (an internship student working on the BAMBI project is a possible candidate). The short-term goal for the team and the PhD student will be to devise a first generic stochastic machine configurable with any generating set. During the same time the source localization and separation problems will be modeled using the Bayesian programming formalism and tested in simulations (using reduced dimensions and simplifying hypotheses). The overall goal of this first year will be to simulate an initial prototype of stochastic machine able to solve simplified versions of our two problems of interest (localization and separation).

During the second year, we will hire a Post-Doctoral researcher to write the extended Bayesian programs corresponding to our two problems. The task of the Post-Doc will be to adapt Bayesian models so that the corresponding Bayesian inference could run on the stochastic machines. This time, we expect to design programs which will be realistic in terms of uncertainty and acoustic constraints. We will define and implement a first architecture for the machine on an FPGA. Finally we will study how to select the generating sets which,

combined, could explore efficiently the probabilistic landscape and find solutions to our two inference problems. The Post-Doc will be under the supervision of L. Girin and will collaborate with E. Mazer to develop the models with the Bayesian programming formalism and with L. Fesquet to evaluate their suitability to run on the proposed architecture.

During the third year we will deliver and test working stochastic machines on the selected problems. All the components of the project will be finalized: Bayesian programs for acoustic models, inference algorithms based on generating sets, hardware architectures able to handle bitwise probabilistic searches and compilers to transform acoustic inference Bayesian programs into the targeted architectures. The team will bring all the components together to build a real stochastic machine and run evaluations on realistic test beds. The Post-Doc researcher will have the responsibility of establishing the experimental set-up as well as the quantitative evaluation of the results in term of accuracy of the predictions made by the machine. The last year of the project will also be devoted to the dissemination of the results, including the PhD student writing the manuscript of his/her thesis and members of the project preparing the workshop and the conference concluding the project.

	T0	T0+6	T0+12	T0+18	T0+24	T0+30	T0+36
<u>Milestones and deliverables</u>	Hiring the <u>Phd</u>		Initial Evaluation in simulation+ hiring of the <u>postdoc</u>	publication in a conference	first evaluation on full scale problem	publication in a journal	<u>Phd Thesis + demonstration of stochastic machines operating on a set of microphones</u>
M1		Generic architecture bitwise sampling and generating sets		Implementation on FPGA and connection to massive random sources			
M2		mathematical formalization	algorithm based on a single generating set	General search strategies based on generating sets for acoustic problems		Optimizing the proposed algorithms (best combination of generating sets)	
M3		Bayesian program for simplified acoustic model and inference		Bayesian programs for real world acoustic models and inferences			
M4			Evaluations in simulation		Evaluations on an FPGA	Experimentation, tuning and evaluation on FPGAs connected to a test bed	

Figure 3: Gantt chart for the project

7 Requested funds

As mentioned above, we plan to involve one PhD student (3 years) and one post-doctoral researcher (2 years) in this project. Both will be funded by PERSYVAL.

Furthermore we plan to fund small hardware for the developments and for the demonstrators (computers, FPGA boards), for a total of 20 k€.

The budget for missions is 50 k€ (15 k€ to accompany the post-doctoral student and the PhD student funded by PERSYVAL and 20 k€ for permanent action members). The funding for permanent members will include an invitation of foreign researchers on a small duration.

We deem important to disseminate the scientific benefits of the project. Therefore we will organize a summer school, e.g. in Aussois, in the long tradition of summer schools organized by INRIA, GIPSA-lab and the Grenoble mathematical community, among others (budget 5 k€). This event will also allow to underline the structuring efforts allowed by the 3 years of the action team.

The total budget is 275 k€ distributed according to the following Table.

One 3-year PhD grant	100k€
One 24-month post-doc position	100k€
Equipment	20k€
Missions	50k€
Workshop organization or summer school	5k€
Total	275k€

8 Annex A: Explicit response to the initial committee

The connection of stochastic machines with event-based sensors (action team e-BaCCuSS) is of a major interest to the future development of stochastic computing. Our opinion is that this connection must be investigated since event-based sensors could directly provide inputs to our stochastic computing device without additional hardware, e.g., by setting the right thresholds in level-crossing sampling scheme. For this reason we believe the two projects are synergic since they reinforce each other by providing new means to process the information coming out of event-based sensors and by providing stochastic inputs to our machine. The otherwise mentioned project HPES is ending and as such will not interfere with MICROBAYES

9 Annex B: PhD Thesis proposal

Title: Design of stochastic machines for source localization and separation

Supervisor: Emmanuel Mazer

Co-supervisors: Didier Piau, Laurent Fesquet

Place: Laboratoire d'Informatique de Grenoble (LIG)

Starting: Fall 2016

The Persyval Labex is offering a PhD grant for a 3-year period. The overall goal of the proposed research is to design, build and program new types of information processing machines and to demonstrate their effectiveness in low-level sensor signal processing such as sound source localization and separation.

The goal of the research is to replace the Von Neuman architecture and the floating point arithmetic units found in CPU and GPU with innovative designs of processing random bit stream. The objective is to reduce the power consumption of computing devices while increasing their processing speed and capabilities. The main part of the work will take place at the LIG, but to address this multi-disciplinary subject the candidate will have the opportunity to collaborate with three other laboratories that are members of Persyval (Institut Fourier, TIMA, and GIPSA-lab).

To get started the candidate will have to improve the design of an existing stochastic machine based on Gibbs sampling and developed at the LIG. The goal is to make the random walk used by Gibbs sampler generic by using generating sets. To do so the candidate will have to get familiar with stochastic computing and Bayesian Programming (LIG), and collaborate with the Institut Fourier on Generating set for random exploration by the Gibbs algorithm and with the TIMA on circuit synthesis. Based on this first improved design the candidate will synthesize an initial prototype on FPGA and will have to devise ways to process the problem of interest: sound source localization and separation, based on algorithms developed at GIPSA-lab. Initial tests will be performed on simplified versions of the problem. Then new designs and programming capabilities will have to be devised to take into account the complexity of the task.

Finally and in collaboration with a post-doc hired on the project, the candidate will test and evaluate the proposed architecture on a real testbed.

10 Annex C: Post-Doc proposal

Title: Implementation and evaluation of stochastic machines for sound source localization and separation

Supervisor: Laurent Girin

Co-supervisor: Emmanuel Mazer

Place: GIPSA-lab and Laboratoire d'Informatique de Grenoble (LIG)

Starting: Fall 2017

The Persyval Labex is offering a 2-year Post-Doctoral position to evaluate the performances of stochastic machines in low-level sensor signal processing such as sound source

localization and separation.

The overall goal is to test non Von Neumann architecture on difficult inference problems related to audio processing. The candidate should have a good prior knowledge about the Bayesian methods used to model and solve the source localization and the source separation problems.

A first objective will be to program simplified versions of these methods using the Bayesian programming formalism. These programs will be tested in simulation and on an initial version of the stochastic machine built during the first year of the project. Then the main task will be to adapt the methods designed to process the acoustic signals to the stochastic algorithms used in the Gibbs samplers, and reciprocally.

After setting an experimental testbed, tests will be conducted to evaluate the proposed architectures, and compare them with more standard method running on conventional machines.