PROCEEDINGS OF SCIENCE

Classification of $\Upsilon(5S)$ Decays with Self-Organizing Neural Networks

Matthias Ullrich[*,†], **Wolfgang Kühn, Sören Lange and Marcel Werner**
for the Belle Collaboration

Justus-Liebig-Universität Giessen
II. Physikalisches Institut
Heinrich-Buff-Ring 16
D-35392 Giessen

The applicability of self-organizing neural networks in order to distinguish between the dominating decay modes of the $\Upsilon(5S)$ resonance has been examined. Variables that are characteristic for each decay mode have been identified and are used as input parameters for the network. The network has been trained with data collected by the BELLE detector at the KEKB energy-asymmetric e^+e^- collider.
The initial conditions of the neural network were fixed using a statistical method. Regions which correspond to different $\Upsilon(5S)$ decay modes have been defined on the final map after training. The optimal rates concerning the largest product of detection efficiency $P_{eff.}$ and correct identification efficiency $P_{cor.}$ have been calculated.

XLIX International Winter Meeting on Nuclear Physics
24-28 January 2011
BORMIO, Italy

[*]Speaker.
[†]E-mail: Matthias.Ullrich@physik.uni-giessen.de

 http://pos.sissa.it/

PoS(Bormio 2011)015

1. Introduction

The Belle experiment (described elsewhere [1]) recorded $\approx 120\ \mathrm{fb}^{-1}$ at a center of mass energy corresponding to the mass of the Υ(5S) resonance. In contrast to the Υ(4S) resonance which decays dominantly into $B\overline{B}$ pairs [2] it can decay due to the higher mass into various final states, as, for example, in excited B mesons (e.g. B^*) or B mesons containing an s-quark (e.g. B_s^*) instead of an u- or d-quark [2]. By analyzing this data sample one should be able to explore a lot of new aspects of beauty dynamics.

When performing an analysis it is crucial to know, which type of B mesons (B_d, B_U, B_S, $B^*_{u,d,s}$) the resonance was decaying into. The B mesons have to be reconstructed exclusively by combining all their decay products, which is a non trivial task.

Here a different approach is investigated: An attempt was made to classify the decay pattern of the Υ(5S) resonance with a self-organizing neural network. This type of artificial neural network was first described by Kohonen and is referred to as Kohonen map [3]. Using these maps to cause different parts of the network to respond in a similar way to certain input patterns is the main goal [3].

Neural networks are able to recognize the main feature of high-dimensional input patterns and to project them into a lower dimensional space whereby the topological properties of the input space are preserved.

Thus, maps can be used for extracting and ordering features of (high-dimensional) data samples.

The major benefit of a Kohonen map is the unsupervised learning process: There is no need for simulated data with a priori known structures and dynamics.

2. The Υ(5S) Resonance

The Υ(5S) state is composed of a b and a $\bar{b}$ quark and has the quantum numbers $J^{PC}=1^{--}$. Hence, it can be produced directly via a virtual photon arising from annihilation of an electron and an positron.

The mass difference to the next less radially excited state Υ(4S) is about 270 MeV. The Υ(5S) mass' is $m_{\Upsilon(5S)}$=(10.865 ± 0.008) GeV [2]. For example, it can decay into excited B meson pairs ($B^*\overline{B}^*$ or $B^*\overline{B}$) or even into excited B mesons containing a strange quark ($B_s\overline{B}_s$, $B_s^*\overline{B_s}$ and $B_s^*\overline{B_s}^*$) instead of a d- or u quark.

Three-body decays ($B\bar{B}\pi$), four-body decays ($B\bar{B}\pi\pi$) and decays into less radially excited Υ states have been seen, too [2]. The branching fraction of some individual decay modes can be found in table 1, more detailed information are available elsewhere [2] [4] [5].

Υ(5S) decay modes	Fraction (Γ_i/Γ)	CL	p (MeV/c)
$B\overline{B}X$	(59 ± 14) %		-
$B\overline{B}$	$(5.5^{+1.0}_{-0.9} \pm 0,4)$ %		1280
$B^*\overline{B} + c.c.$	(13.7 ± 1.3 ± 1.1) %		-
$B^*\overline{B}^*$	$(37.5\ ^{+2.1}_{-1.9} \pm 3.0)$ %		-
$B^{(*)}\overline{B}^{(*)}\pi$	< 19.7 %	90 %	-
$B\overline{B}\pi\pi$	< 8.9 %	90 %	441
$B_s^{(*)}\overline{B}_s^{(*)}$	(19.3 ± 2.9) %		-
$\Upsilon(1S)\pi^+\pi^-$	$(5.3 \pm 0.6) \times 10^{-3}$		1288
$\Upsilon(2S)\pi^+\pi^-$	$(7.8 \pm 1.3) \times 10^{-3}$		763
$\Upsilon(3S)\pi^+\pi^-$	$(4.8^{+1.9}_{-1.7}) \times 10^{-3}$		416
$\Upsilon(1S)K^+K^-$	$(6.1 \pm 1.8) \times 10^{-4}$		933

Table 1: Selected Υ(5S) decay modes ([2] [4]). P is the maximum momentum of the decay products, CL the confidence level. The sum over the branching fraction does not have to be equal to one because decay modes are submodes of others. The excited states decay into their ground states via emission of a γ. Here B denotes a B^+ or B^0 meson and $\overline{B}$ a $\overline{B}^-$ or $\overline{B}^0$ meson. A π stands for π^+, π^- or π^0.

3. Self-Organizing Neural Networks

A self-organizing map (SOM) is an artificial neural network (ANN) which learns in an unsupervised mode[3]. Using so-called neighborhood functions it preserves topological properties of the input space. This characteristic predestines the networks to visualize low-dimensional views of high-dimensional data.
It operates like most ANN in two modes: training and mapping (of identified target regions). Training (or learning) builds the map using input examples; mapping automatically classifies an input vector.

3.1 The Basic SOM Algorithm

Let $x = [\xi_1, \xi_2, \dots, \xi_n] \in \mathscr{R}^n$ be an arbitrary vector. This vector x may be compared with all the m_i in any metric and the best-matching node (i.e. the smallest distance of both vectors), signified by the subscript w, is determined:

$$w = \underset{i}{\operatorname{argmin}}\ ||x - m_i|| \qquad \Leftrightarrow$$
$$||x - m_w|| = \min_i ||x - m_i|| \ . \qquad (3.1)$$

Nodes which are topographically close to the winner node m_w in the node-space (up to certain geometric distance) will activate each other to learn from the same input x. This will lead to a local relaxation effect on the weight vectors of the neurons in this neighborhood. The learning process of the standard SOM algorithm can be formulated as follows:

$$m_i(t+1) = m_i(t) + h_{wi}(t)[x(t) - m_i(t)] \ ,$$

where $t = 0, 1, 2, \ldots$ is integer and a discrete time coordinate. The initial values $m_i(0)$ may be arbitrary. The function $h_{wi}(t)$ plays a crucial role in the relaxation process, acting as the so-called neighborhood function.
The behavior of this function is crucial for global convergence of the map. It is necessary that

$$h_{wi}(t) \rightarrow 0 \text{ when } t \rightarrow \infty \,.$$

Usually the neighborhood function depends also on the distance between node w and node i, i.e $h_{wi}(t) = h_{wi}(||r_w - r_i||, t)$, where r_w and r_i are the position vectors of nodes in node space, respectively.
Two simple choices of neighborhood functions appear frequently in literature [3]. The simpler one refers to a neighborhood set of array points around the winner node w. The second widely applied kernel is a much smoother one: $h_{wi} = \alpha(t) \cdot exp\left(-\frac{||r_w - r_i||^2}{2\sigma^2(t)}\right)$, with $\alpha(t)$ the so-called learning rate and $\sigma(t)$ the width of the kernel. Either $\alpha(t)$ or $\sigma(t)$ has to be decreasing in time.

3.2 Quality Assessment

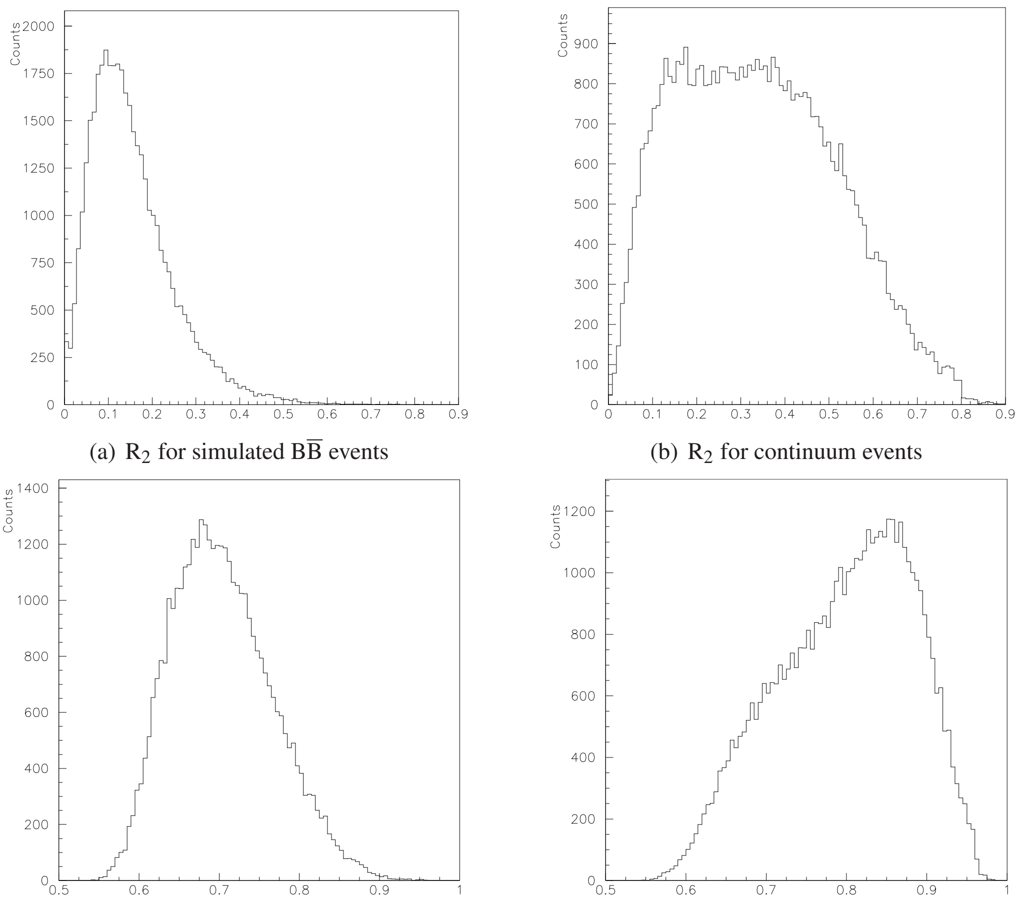
Considering a perfect training (learning) process, each node i will represent a certain volume of the input space. If data out of this volume is presented to the network the corresponding node will be the winner. For the ideal case each node should respond equal times (concerning the input data) [6].
Plotting the winner rate as a function of each node i the distribution should be flat.
$\chi^2 = \sum_i \frac{(H_i - \overline{H})^2}{\overline{H}^2}$ is a gauge for the derivation of this uniform distribution. H_i is the winner rate of node i and $\overline{H}$ is the average value of all H_i. The smaller χ^2, the better the quality of convergence.

4. Variables for the Identification of Different Υ(5S) Decay Modes

It is crucial to find variables which may be appropriate to characterize and so be able to separate the different dominating Υ(5S) decay modes. These variables will be later used as input data for the self-organizing neural network.
A large part of the data are continuum events, e.g. QED events (Bhabha scattering and Bhabha radiative processes) which should be rejected.
The data used for this analysis is already preselected by a skimming criteria sensitive to more than 99% of all produced hadrons. After skimming the expected ratio of Υ(5S) events to continuum events is $\frac{1}{3.5}$ [7]. Hence, variables for the separation of continuum events and Υ(5S) events are important.
A number of variables have been developed for the means of continuum suppression. They all take advantage of the differences in event shapes of continuum and $\mathrm{B\overline{B}}$ events. Resonant $\mathrm{B\overline{B}}$ events release only little energy and tend to be spherical. In contrast a large energy release indicates a continuum event. The q and $\bar{q}$ jets of continuum events tend also to emerge collimated back to back [8].This analysis uses two variables for event shape characterization: R_2 and thrust.

R_2 is the ratio of the second and the zeroth Fox-Wolfram moment. The Fox-Wolfram moments H_l are defined by $H_l = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{E^2_{\mathrm{vis}}} P_l(cos(\theta_{ij}))$ [9], where θ_{ij} is the opening angle between hadrons

i and j, and $\vec{p}$ the corresponding momentum. P_l are the Legendre polynomials and E_{vis} is the total visible energy of an event. A value of R_2 closer to zero indicates more spherical events, whereas lager values of R_2 indicate continuum events (compare to figure 1).
The so-called *thrust* measures the alignment of particles in an event along a common axis. It is defined by $T = \max_{|\vec{n}|=1} \frac{\sum_i |\vec{n}\cdot\vec{p}_i|}{\sum_i |\vec{p}_i|}$ [10], where $\vec{p}$ is the momentum and $\vec{n}$ is the normalized vector in which the maximum alignment is found. The higher the thrust, the more jet-like the event is (compare to figure 1).



(a) R_2 for simulated $B\overline{B}$ events

(b) R_2 for continuum events

(c) Thrust for simulated $B\overline{B}$ events

(d) Thrust for continuum events

Figure 1: The distributions of R_2 in (a) and (b) and the thrust in (c) and (d) show that these variables are useful for the separation of continuum and $B\overline{B}$ events.

The *visible energy* $E_{\text{vis.}}$ (the sum of the energy of all particles measured by the detector in one collision) is appropriate to differentiate between continuum and $B\overline{B}$ events [11], and is used as input parameter, too.
The number of charged pions in restricted momentum intervals proves as a good indicator for the many-body decays $\Upsilon(5S) \to B\overline{B}\pi$ or $\Upsilon(5S) \to B\overline{B}\pi\pi$ respectively. The number of pions is enhanced compared to Υ(5S) decays which decay only into two B mesons.
The optimal range for the momentum interval for $\pi^{\pm}$ arising from the 4-body mode is

$$0.03 \text{ GeV} < |\vec{p}_{\text{cms}}| < 0.1 \text{ GeV}$$

and the interval for the 3-body decay mode is

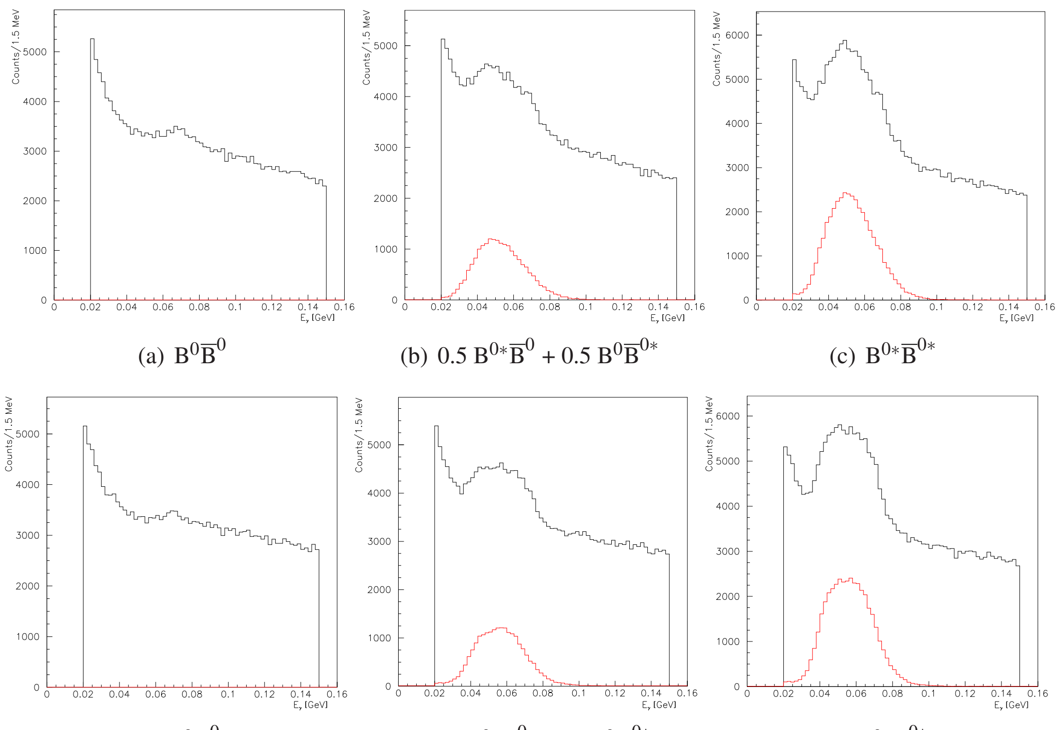
$$0.1 \text{ GeV} < |\vec{p}_{\text{cms}}| < 0.25 \text{ GeV};$$

compare to [11].

The pions emerging directly from Υ(5S) decays have another feature which allows to distinguish them from $\pi^{\pm}$ coming from B decays. Their production vertex is the decay vertex of the Υ(5S) resonance which is in a distance of less than 100μm from the interaction point and thus can be approximated by z=0.

Thus, the number of pions around the interaction point (-0.1 cm < IP_z < 0.125 cm, -0.05 cm <$IP_{x,y}$<0.05 cm) is chosen as input parameter for the neural network (compare also to [11]).

The excited B mesons (B^*, B_s^*) decay dominantly into Bγ or $B_s\gamma$. Hence, the photons may be useful to identify Υ(5S) decays into excited B mesons. Again, the number of photons in a certain energy interval is used to identify Υ(5S) decays involving excited B mesons. In [11] is shown that the number of photons in the energy range 0.03 GeV < E_γ < 0.07 GeV differ for decays involving excited B mesons and such decays which do not (compare also to figure 2). More revealing variables for the mean of identification of excited B mesons have not been found.



(a) $B^0\overline{B}^0$ (b) 0.5 $B^{0*}\overline{B}^0$ + 0.5 $B^0\overline{B}^{0*}$ (c) $B^{0*}\overline{B}^{0*}$

(d) $B_s^0\overline{B}_s^0$ (e) 0.5 $B_s^{0*}\overline{B}_s^0$ + 0.5 $B_s^0\overline{B}_s^{0*}$ (f) $B_s^{0*}\overline{B}_s^{0*}$

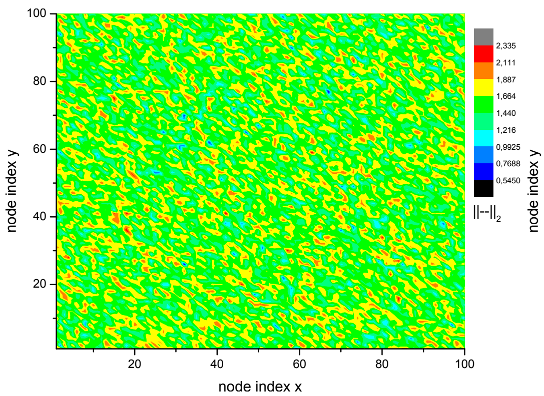
Figure 2: **Black curves**: Simulated energy distribution of the photons for different Υ(5S) decay modes in GeV. The **red curves** shows the energy distributions of the photons coming directly from the respective excited B mesons.

The underlying idea for the identification of B_s mesons was to use the strangeness which is in contrast to $B_{u,d}$ mesons already available in the initial state. The first feasible idea was just to

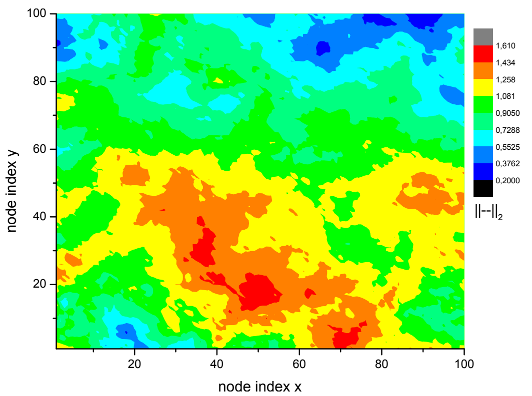
count the number of Kaons in the final state — they are the "carrier" of the strangeness. However a detailed analysis showed, that this number is not really useful (compare to [11]).
As a variable which is sensitive to the strangeness the number of Φ(1020) decaying into K^+K^- has been used finally — the amount of Φ(1020) differs significantly in B and B_s decays [2] [11].
More "basic" variables sutiable for the purpose of separation of Υ(5S) decays have not been found (compare to [11]).

5. Trainings Data, Self-Organizing Map and Parameter Optimization

1.4 million events (corresponding to 0.34 % of the total Υ(5S) data set) collected by the Belle detector at a center of mass energy of 10.871 GeV have been used for training.
The map consist of 10 000 units arranged in a 100×100 grid; an 8-dimensional vector is associated to each unit. Each component of the nodes' weight vectors has been initialized with random numbers.
Following functions have been chosen for the time dependent learning rate α and width σ of the neighborhood: $\alpha(t) = \alpha_{in}(1-(\alpha_{in}-\alpha_f)\cdot\frac{t}{t_{max}})$ and $\sigma(t) = \sigma_{in}(1-(\sigma_{in}-\sigma_f)\cdot\frac{t}{t_{max}})$. The initial parameters have been determined optimal by the χ^2 method (compare to 3.2).
Weighting of the components of the input vectors is done by multiplying each component by a certain factor. This might have a positive impact on the final results [6]. The optimal weights have again been determined by the χ^2 method. Eventually, the number of pions in the momentum interval from 0.03 till 0.1 GeV is weighted by a factor of 2, whereas all other input parameters have the weight 1. The optimal values for the learning rate and the width of the neighborhood function have been determined to be $\alpha_{in} = 1$, α_f=0.1, σ_{in}=4 and σ_f =3.
Figure 3 shows a measure of the nodes' weightvectors before and after training: The smoother map behaviour after training is obvious.



(a)



(b)

Figure 3: **(a)** shows the contour plot of the initial and **(b)** of the trained map. The x- and y-coordinates refer to the position of the nodes in node space. The color indicates the euclidean norm of the corresponding weight vectors. The smoother map behaviour of **(b)** is evident.

6. Results

To evaluate the map after training, Monte-Carlo data with a priori known decay processes of the $\Upsilon(5S)$ resonance are required. Therefore Monte-Carlo data was produced using EvtGen (a Monte-Carlo generator for B-physics) and a full detector simulation. The map has been calibrated with this data and decays containing B mesons yield to responses in spatially limited regions (compare to figure 4 and [11]). The calibration is essentially to count how often a unit matches best to input vectors from the Monte-Carlo samples.

Eventually, regions which can be assigned to certain decay channels have been restricted by linear functions (splines). An event within the restricted area will be counted as the dedicated decay process. Within the first step the functions which limit the regions were set manually, but an additional parameter b is introduced which allows to vary the different functions and as a consequence the area of the regions.

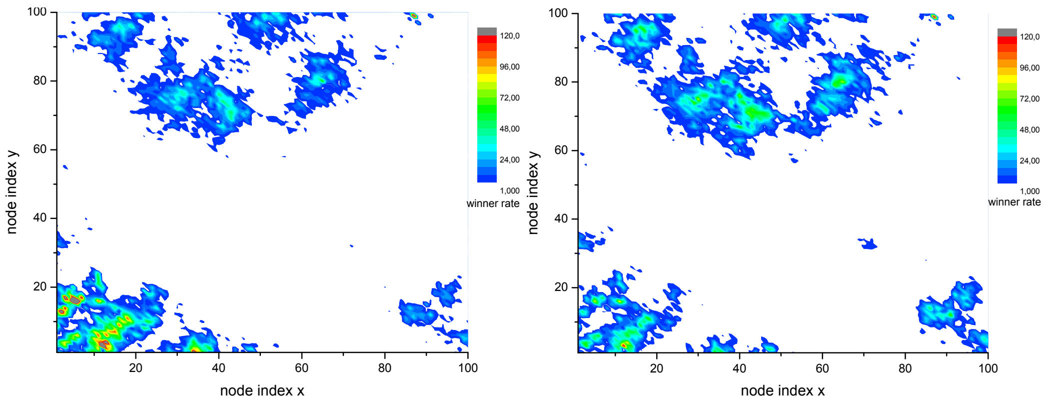
The parameter b is optimized relying on the quality parameter $P_{cor.}$, the probability that an event is identified correctly and $P_{eff.}$, the detection efficiency of a certain decay channel of each decay mode.

The optimal rates concerning the largest product of $P_{cor.}$ and $P_{eff.}$ for the different decay modes are:

$$\begin{array}{rll} B\overline{B}\pi\pi: & P_{eff.} = (31.7 \pm 0.1)\% & P_{cor.} = (8.26 \pm 0.03)\% \\ B_s^{(*)}\overline{B}_s^{(*)}: & P_{eff.} = (35 \pm 18)\% & P_{cor.} = (5.4 \pm 2.8)\% \\ B^{(*)}\overline{B}^{(*)}: & P_{eff.} = (80.78 \pm 0.74)\% & P_{cor.} = (25.7 \pm 0.23)\% \\ B^{(*)}\overline{B}^{(*)}\text{ inclusive}: & P_{eff.} = (83.30 \pm 0.69)\% & P_{cor.} = (27.72 \pm 0.23)\% \\ B_{(s)}^{(*)}\overline{B}_{(s)}^{(*)}\text{ inclusive}: & P_{eff.} = (81.6 \pm 1.1)\% & P_{cor.} = (35.61 \pm 0.49)\% \end{array}$$

Here B denotes $B_{u,d}$ mesons with charge +,- or 0. π stands for the charged as well as for the uncharged π mesons. The errors of these quantities are given as standard deviations.

Plots, information and details about the relationship of all the different parameters and functions are available in [11].



(a) $\Upsilon(5S)\rightarrow B^0\overline{B}^0\pi^+\pi^-$ (b) $\Upsilon(5S)\rightarrow B^0\overline{B}^0$

Figure 4: The x- and the y-axes indicate the position of the neurons in node space. The color indicates the winner rate of each neuron. Events which contain pions coming directly from the $\Upsilon(5S)$ decays tend to show up with an increased rate in the left bottom corner.

7. Summary and Conclusion

The applicability of self-organizing neural networks in order to classify the decay patterns of the Υ(5S) resonance has been examined. The dominating decay channels, in particular $B\overline{B}\pi\pi$, $B\overline{B}\pi$, $B\overline{B}$, $B_s\overline{B}_s$ as well as the respective excited modes (e.g. $B^*\overline{B}$) should be distinguished.
Variables which characterize the respective decay processes are determined and optimal initial parameters of the neural network were calculated. Regions which are typical for the particular decay modes have been worked out and optimized.
A clear identification of the individual decay modes is not possible (using self-organizing neural networks). The decays are too similar to be distinguished by "basic" variables only.
However, it is possible to enrich the fraction of certain decay channels. As an example may hold the decay mode "$B^{(*)}_{(s)}\overline{B}^{(*)}_{(s)}$ *inclusive*": In (3.20 ± 0.03) of 9 events B mesons can be found now, while originally only 2 of 9 events contained $\Upsilon(5S)$ decays (compare to [7]). Thus, the number of B mesons could be increased by a factor of 1.6.

References

[1] A. Abashian *et al.* (Belle Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **479**, 117 (2002)

[2] C. Amsler *et al.* (Particle Data Group), Particle Physics Booklet, July 2008

[3] Springer Series in Information Sciences, T. Kohonen, Self-Organizing Maps, First Edition, Springer-Verlag Berlin Heidelberg

[4] A. Drutskoy *et al.* (BELLE collaboration), Phys. Rev. D 81, 112003 (2010), arXiv:1003.5885v3 [hep-ex]

[5] R. Louvot *et al.* (BELLE collaboration), Phys. Rev. Lett. 102, 021801 (2009), arXiv:1003.5312v1 [hep-ex]

[6] J. S. Lange, Extraktion von Bremsstrahlungsereignissen in Proton-Proton-Reaktionen durch Anwendung künstlich neuronaler Netzte, Dissertation, Ruhr-Universität Bochum 1994

[7] A. Schwartz, CKM'06, Belle physics results at the Y(5S), Nagoya Daigaku, December 14, 2006

[8] BaBar Glossary, July 27, 2010, http://www.slac.stanford.edu/babar-pro-new/search.pl?letter=Entire+Glossary

[9] G.C. Fox and S. Wolfram, Nucl. Phys. B149 (1979) 413

[10] S. Brandt, Ch. Peyrou, R. Sosnowski and A. Wroblewski, Phys. Lett. 12 (1964) 57

[11] M. Ullrich, Classification of Υ(5S) Decays with Self-Organizing Maps, Master Thesis, Universität Giessen 2010