

# Method for High Accuracy Multiplicity Correlation Measurements

C. Søgaard<sup>1,a</sup> and K. Gulbrandsen<sup>2,b</sup>

<sup>1</sup>Lund University, Sweden

<sup>2</sup>Niels Bohr Institute, Discovery Center, Copenhagen, Denmark

**Abstract.** A novel method for measuring particle multiplicity correlations is presented. The method takes reduced acceptance and detection efficiency into account, in a purely statistical manner. The method does therefore not require any information from simulations - only knowledge of the detectors capability. The method provides the ability to measure forward-backward particle multiplicity correlations with high accuracy and negligible bias. The validity of the method is provided through numerous simulations. The correlation values calculated from carefully selected detection acceptances and efficiencies are compared to the correlation values at the event-generator level. Further validity to the method is given by incorporating different event generators into the simulations. Although the emphasis is on fwd-bwd correlations, the method can be extended to other multiplicity correlation measurements.

## 1 Introduction

Particle correlations play an important role in understanding the underlying event of high energy hadronic collisions. Discussions on the physics of the underlying event is for example found in [1–3]. The method presented here provides means for measuring particle multiplicity correlations, using detection systems with reduced acceptance and/or reduced detection efficiency. Specifically the measurement of forward-backward correlations is addressed.

The most common method for measuring forward-backward correlations, with less than ideal detection capability, is to correct the directly observed correlation factors  $b_{\text{experiment}}$  using a correcting factor  $c$  derived from realistic simulations. The correction factor is found by

$$c = \frac{b_{\text{primary}}}{b_{\text{simulation response}}} \quad (1)$$

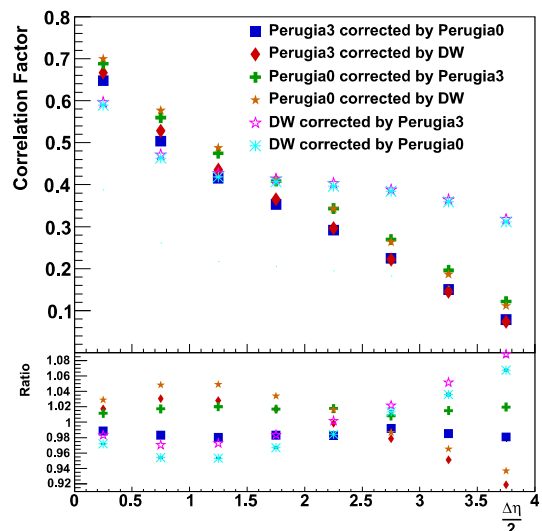
where  $b_{\text{primary}}$  is the correlation factors obtained from the particle distribution at the generator level, and  $b_{\text{simulation response}}$  are the correlation factors of the particle distribution from the simulated detector response. The corrected correlation factors  $b_{\text{final}}$  are then calculated by

$$b_{\text{final}} = c \cdot b_{\text{experiment}} \quad (2)$$

This method may be inaccurate on at least two accounts. The method requires accurate description of the detector in the simulation. Although this can be achieved, high enough accuracy may be cumbersome and time consuming. Even disregarding this fact, the method is by construction biased. The correlation factors  $b_{\text{primary}}$  and

$b_{\text{simulation response}}$  of equation (1) both depend on the event generator chosen for the simulation. The bias of this simple correction method stems from these differences.

An attempt to illustrate the bias of the described simple method is given in figure 1. Since no data from experiment is available for this paper, the effect has been illustrated by performing simulations with three different



**Figure 1.** Correlation factors obtained when only 60% of the total acceptance exists. The situation is drawn in figure 3. The corrections have been performed using correction factors from simulations with different event-generators. The bottom part of each figure shows the ratio of the corrected correlation factor to the actual primary correlation factor for the tune that the data was generated from.

<sup>a</sup>e-mail: carsten.sogaard@hep.lu.se

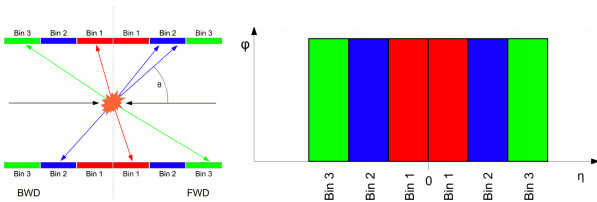
<sup>b</sup>e-mail: gulbrand@nbi.dk

event generators. The event generators are all Pythia6 [4] but with different settings, called tunes. The tunes in the examples are: Perugia3, Perugia0, and DW [5]. Three correction factors  $c_{\text{Perugia3}}$ ,  $c_{\text{Perugia0}}$  and  $c_{\text{DW}}$  have been calculated using (1). The correlation factors  $b_{\text{experiment}}$  are replaced by  $b_{\text{simulation response}}$  from a simulation with another event-generator. In this manner, each simulated detector response is corrected by two correction factors originating from two different event generators. This provides 6 examples as shown in the figure. The discrepancies between the correlation factors of the primary distribution to the corrected correlation factors are shown in the bottom. These discrepancies illustrate the bias of the method. As shown, the biases in these examples have values of up to 8%.

Although the biases may not seem too large, it should be noted that this bias is only one of possible many effects that will alter the final result of an experiment. The method presented later achieves the same measurement but with biases of the order of 1 per mill. Furthermore, the method is based on statistical arguments and calculations. The method does therefore not require any simulations - only knowledge of detection acceptance and efficiency.

## 2 Measuring Forward–Backward Correlations

The geometry of a collision experiment can be divided into two hemispheres separated by the plane perpendicular to the beam axis and intersecting at the collision point. The hemisphere where  $\theta < \pi/2$  is usually termed forward and the hemisphere where  $\theta > \pi/2$  is usually termed backward. This is illustrated in figure 2. Often the coordinate is transformed into the so-called pseudo-rapidity  $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$ . Forward–backward correlations are measured by determining the particle multiplicities of two detection regions with one region located in each hemisphere. Usually the two regions are located equidistantly from  $\eta = 0$  and have equal widths (in  $\eta$ ) while spanning the entire azimuthal angle. In some cases correlations between bins which span less than  $2\pi$  in azimuth are measured [3].



**Figure 2.** Left: Schematic view of the two hemispheres. The dashed line indicates  $\theta = \frac{\pi}{2}$  (or  $\eta = 0$ ). Each hemisphere is then again divided into solid angles, spanning a subset of the polar angle and the full azimuthal angle. Right: The detection regions have been mapped to a two dimensional figure with coordinates in  $\eta$  and  $\varphi$ . Note equal size bins in  $\theta$  on the left do not result in equal size bins in  $\eta$  on the right. Hence, the mapping is not to scale.

With this definition, the multiplicity, in a single event, of a forward region is then  $N_f$ , while the multiplicity of a

backward region becomes  $N_b$ . The correlation factor  $b$  is then calculated as

$$b = \frac{\text{Cov}(N_f, N_b)}{\sqrt{\text{Var}(N_f)\text{Var}(N_b)}} = \frac{\langle N_f N_b \rangle - \langle N_f \rangle \langle N_b \rangle}{\sqrt{(\langle N_f^2 \rangle - \langle N_f \rangle^2)(\langle N_b^2 \rangle - \langle N_b \rangle^2)}} \quad (3)$$

In summary, the correlation factor is calculated from the five moments  $\langle N_f \rangle, \langle N_b \rangle, \langle N_f N_b \rangle, \langle N_f^2 \rangle, \langle N_b^2 \rangle$  of the bivariate distribution of  $(N_f, N_b)$ .

The measurement is therefore quite simple if the detection system used does not obscure the particle distribution.

### 2.1 The Effect of Detection Efficiency

Full hermeticity of a detection system rarely exists. This in turn means that not all particles produced in the collision are detected. Such an effect will alter the bivariate distributions, from which the correlation factors are obtained. This is the topic of the following sections.

The joint probability distribution for the produced particles  $P^P(N_f, N_b)$  contains the information we wish to extract and is normalized such that

$$\sum_{N_f=0}^{\infty} \sum_{N_b=0}^{\infty} P^P(N_f, N_b) = 1 \quad (4)$$

The moment generating function can be constructed for this distribution by

$$\text{mgf}^P(t_f, t_b) = \sum_{N_f=0}^{\infty} \sum_{N_b=0}^{\infty} P^P(N_f, N_b) e^{N_f t_f + N_b t_b} \quad (5)$$

where derivatives with respect to  $t_f$  and  $t_b$  evaluated at  $t_f = 0$  and  $t_b = 0$  produce the moments of the distribution.

Assuming a uniform efficiency  $\varepsilon_f$  over the forward region and  $\varepsilon_b$  over the backward region the moment generating function for the detected distribution becomes

$$\begin{aligned} \text{mgf}^D(t_f, t_b) = & \sum_{N_f=0}^{\infty} \sum_{N_b=0}^{\infty} P^P(N_f, N_b) \left[ \varepsilon_f e^{t_f} + (1 - \varepsilon_f) \right]^{N_f} \\ & \times \left[ \varepsilon_b e^{t_b} + (1 - \varepsilon_b) \right]^{N_b} \end{aligned} \quad (6)$$

The moments of the distribution are calculated and the cumulants are constructed. The observed cumulants then become

$$\langle N_{f,b}^D \rangle = \langle N_{f,b}^P \rangle \varepsilon_{f,b} \quad (7)$$

$$\text{Cov}(N_f^D, N_b^D) = \text{Cov}(N_f^P, N_b^P) \varepsilon_f \varepsilon_b \quad (8)$$

$$\text{Var}(N_{f,b}^D) = \text{Var}(N_{f,b}^P) \varepsilon_{f,b}^2 + \langle N_{f,b}^P \rangle \varepsilon_{f,b} (1 - \varepsilon_{f,b}) \quad (9)$$

Equations (7) and (8) show that these cumulants only depend on the efficiencies of the regions, while equation (9) depends not only on the efficiencies, but also on the average. The latter result is somewhat surprising.

These equations are valid when all particles have the same detection probability.

## 2.2 Accounting for Event-shape

In the case of limited acceptance in the detection region, i.e. the forward or backward region has a 'hole', the results from before do not hold. The distribution of produced particles is not uniform in azimuth within an event. Jets, for example, produce large particle multiplicities in somewhat small solid angles. This non-uniform geometric distribution of particles within an event is what is here termed the event-shape. The detection efficiency then varies between events, depending on the geometric distribution of particles, violating the assumption of a uniform detection efficiency of particles.

The event shape is taken into account by exploiting the fact that over many events, the particle distribution is uniform in the azimuthal angle. By dividing each region into  $m_\varphi$  azimuthal segments, a uniform (or close to) efficiency can be achieved in each segment. The optimal number of segments depends on the actual experiment. Equations 7,

8 and 9 still hold for each segment. To calculate the cumulants of the distribution for the full region from the cumulants obtained from the segments, the following relations are used

$$\langle N_{r_1, r_2} \rangle = \sum_{i=0}^{m_\varphi} \langle N_{(r_1, r_2), i} \rangle \quad (10)$$

$$\text{Cov}(N_{r_1}, N_{r_2}) = \sum_{i=0}^{m_\varphi} \sum_{j=0}^{m_\varphi} \text{Cov}(N_{r_1, i}, N_{r_2, j}) \quad (11)$$

where  $r_1, r_2$  are  $f$  or  $b$ . In the case when  $r_1$  and  $r_2$  are the same the covariance becomes the variance, since

$$\text{Cov}(N_{r_1}, N_{r_1}) = \text{Var}(N_{r_1}) \quad (12)$$

The full derivation of the final result is omitted in this document. However, the full derivation can be found in [6]. The final formula for the covariance is

$$\begin{aligned} \text{Cov}(N_{r_1}^P, N_{r_2}^P) = & m_\varphi \cdot \frac{\sum_{i_\varphi=1}^{m_\varphi} \text{Cov}(N_{r_1, i_\varphi}^D, N_{r_2, i_\varphi}^D)}{\sum_{i_\varphi=1}^{m_\varphi} \varepsilon_{r_1, i_\varphi} \varepsilon_{r_2, i_\varphi}} \\ & + m_\varphi \cdot \sum_{s=1}^{m_\varphi-1} \left\{ \frac{\sum_{i_\varphi=1}^{m_\varphi-s} \text{Cov}(N_{r_1, i_\varphi}^D, N_{r_2, i_\varphi+s}^D) + \sum_{i_\varphi=1}^s \text{Cov}(N_{r_1, m_\varphi+i_\varphi-s}^D, N_{r_2, i_\varphi}^D)}{\sum_{i_\varphi=1}^{m_\varphi-s} \varepsilon_{r_1, i_\varphi} \varepsilon_{r_2, i_\varphi+s} + \sum_{i_\varphi=1}^s \varepsilon_{r_1, m_\varphi+i_\varphi-s} \varepsilon_{r_2, i_\varphi}} \right\} \\ & - \delta_{r_1 r_2} \cdot m_\varphi \cdot \frac{\sum_{i_\varphi=1}^{m_\varphi} \varepsilon_{r_1, i_\varphi} (1 - \varepsilon_{r_1, i_\varphi})}{\sum_{i_\varphi=1}^{m_\varphi} \varepsilon_{r_1, i_\varphi}^2} \cdot \frac{\sum_{i_\varphi=1}^{m_\varphi} \langle N_{r_1, i_\varphi}^D \rangle}{\sum_{i_\varphi=1}^{m_\varphi} \varepsilon_{r_1, i_\varphi}} \end{aligned} \quad (13)$$

The Kronecker delta  $\delta_{r_1 r_2}$  ensures proper calculation of the variance, when the two segments in question are identical.

## 3 Verification of the Method

Several realistic simulations have been performed to verify the proposed method. In all cases the event-generator is Pythia6 with the three different tunes: Perugia3, Perugia0 and DW. The tunes have been chosen for their difference in their distributions of correlation factors, rather than their physics descriptions.

The primary particle distributions, from the generator, have been subjected to geometric acceptance cuts, where particles with vectors towards certain directions have been omitted in the data sample. This mimics a detector, where certain parts are unresponsive, so-called dead areas. In another case the primary particles have been subjected to a probability for detection. This mimics detector with less than perfect detection efficiency.

### 3.1 Limited Acceptance

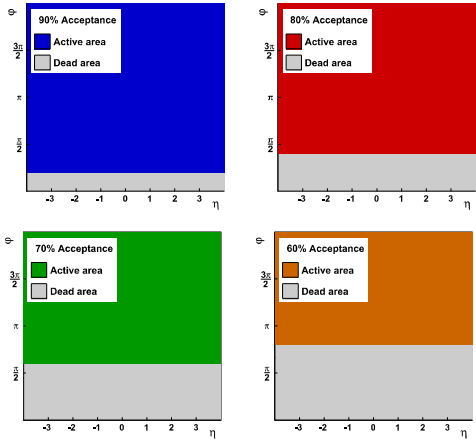
The first example consists of four cases, where the acceptance has been reduced in azimuth for all regions. In

the four cases the geometric acceptances cover 90%, 80%, 70% and 60%. This is depicted in figure 3. The coloured areas represent active areas where particles can be detected. The gray areas represent inactive or 'dead' areas, where particles cannot be detected.

This is a rather extreme case. It has been included, not only to test and verify the method, but also as an attempt to quantify the size of the bias. The reason that the acceptance has not been reduced further is that equation (13) requires that the acceptance in any given region (forward or backward) must be larger than 50%. If this is not the case, the denominator will become zero in at least one term.

The correlation factors have been measured and calculated using the method described earlier. Since the acceptance is exactly known and no other source of noise has been added, the result shows the accuracy of the method. The data sample consists of 5 different runs each consisting of one million events.

The correlation factors are drawn in figure 4. The ratio of the correlation factors of the primary distribution to the correlation factors of the 'detected' distribution is shown in the bottom. This indicates that the bias of the method with these data samples is of the order of a few per mill.



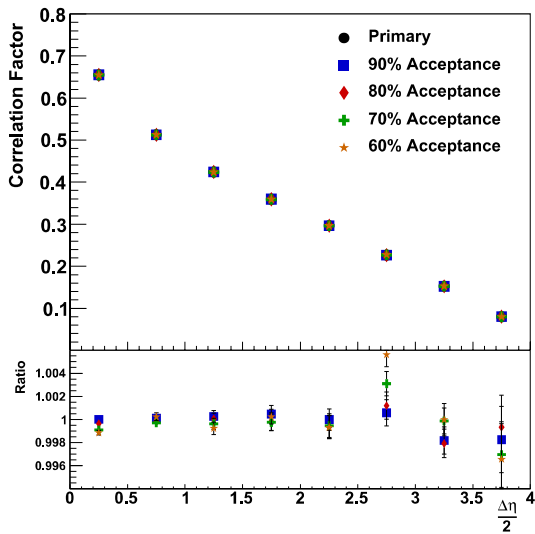
**Figure 3.** The figure shows the geometric acceptances. The acceptances cover 90%, 80%, 70% and 60% of the azimuthal angle respectively. The coloured regions correspond to active areas, which detect particles. The gray areas represent inactive or 'dead' regions where no particle detection takes place.

This demonstrates that, even in this extreme case, the bias of the method is small and can most likely be neglected if applied to experimental data.

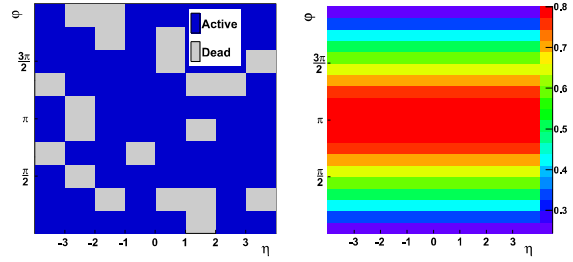
### 3.2 Random 'Holes'

In this example the dead areas are randomly distributed over the region of interest, as would be with 'holes' in the detector. This is a more realistic case compared to the case before. The distribution of holes is shown in figure 5.

In this case the results for different choices of segmentation in azimuth is shown in figure 6. The effect of in-



**Figure 4.** The measured correlation factors, using this method, with acceptances varying from 90% to 60%. In the bottom, the ratios of the measured to primary correlation factors. Note that the resulting bias is at the permil level. The depicted errors are statistical.



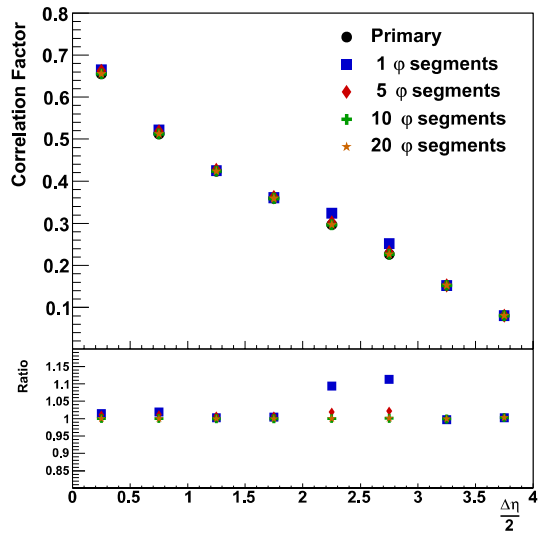
**Figure 5.** Left: The dead areas of the 'detector' are randomly distributed as holes. No bin has less than 60% acceptance. Right: The efficiency along  $\phi$ . In both cases the efficiency is defined in the range  $0.2 < \varepsilon < 0.8$  and are smooth sine functions. The non-continuous appearance of the efficiencies is a result of binning.

creased segmentation is largest for the points at  $\eta = 2.25$  and  $\eta = 2.75$ . Higher accuracy is achieved until a segmentation of 10 segments is reached. Further segmentation does not improve, since at 10 segments, the holes are contained within a segment. It can be shown, analytically, that the 10 and 20 segment cases, in this particular example, yield identical results. This agrees with the argument that smaller segments result in efficiencies which vary less. In the two cases with 10 and 20 segments respectively, the efficiencies are either 1 or 0 in each segment.

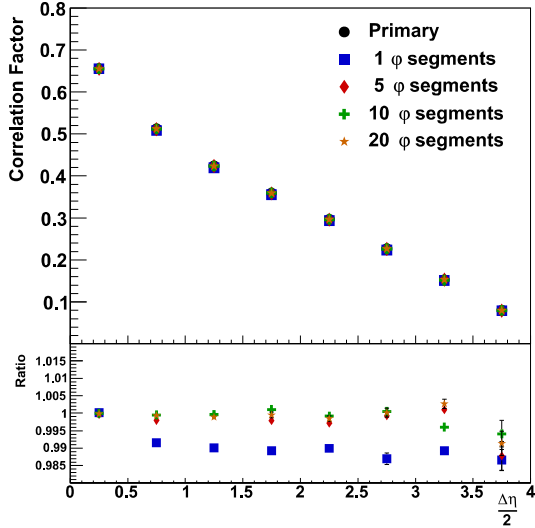
### 3.3 Efficiency Gradient

The final example involves varying efficiencies. The efficiency varies in azimuth and the efficiency map is depicted in figure 5.

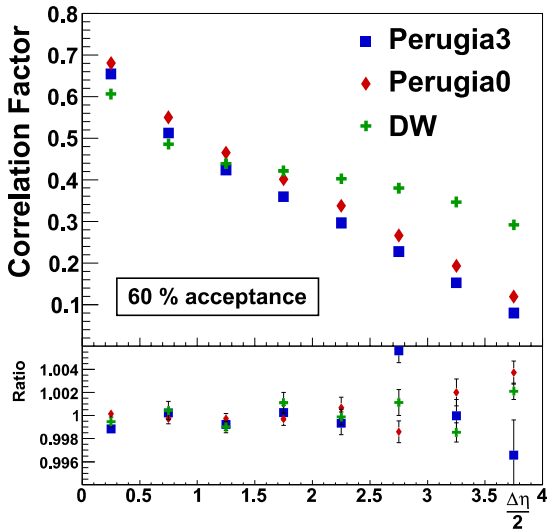
The efficiency gradient is a smooth sine function with values, in the detection region, in the range  $0.2 < \varepsilon < 0.8$ . The figures show a non-continuous efficiency gradient, but



**Figure 6.** The correlation factors obtained in a detector with the 'holes' as shown in figure 5. Again the discrepancies are of the order of a percent, if the segmentation is small enough. More than 10 segments will not improve the result, since 10 segments ensures that all the holes are contained within a segment.



**Figure 7.** The correlation factors when the primary particles have been subjected to a detection probability. Although the detection efficiency becomes as low as 20%, the bias of the method is still less than a percent, if one chooses enough segments.



**Figure 8.** Comparison of correlation factors obtained from the output of the three tunes. The example with an azimuthal acceptance of 60%, shown in figure 3, is applied. The results show that the method exhibits no dependence on the event generator.

this is a result of the binning. The figure then shows the average efficiencies measured in each segment. This efficiency is used in the equation (13), while the detection probability follows the smooth function mentioned earlier.

The correlation factors have been calculated using the same formula as in the previous cases. The difference is that this time, the efficiency will not become uniform, no matter how small the segments are chosen.

However, making the segments smaller does improve the result. This is seen in figure 7 where the results are drawn. Although the efficiencies decrease to 20% the

method still obtains results with discrepancies less than a percent, if small enough segments are chosen.

### 3.4 Comparison between Event Generators

Finally the dependence of the method on event generators is tested. In this case the example with 60% azimuthal acceptance from figure 3 is applied to samples from the three different tunes. The correlation factors are obtained through the method discussed here. The results are shown in figure 8

Although the discrepancies are not identical between the tunes, the discrepancies show no tendency in any of the tunes. The conclusion drawn from this observation, is that there is not bias dependence of the method on the event generator used in this example.

## 4 Summary

We have presented a method for measuring multiplicity correlations with high accuracy using detection systems with less than perfect efficiency and/or acceptance. The method assumes no specific properties of the distribution and is founded entirely on statistical arguments.

The method has been verified through numerous simulations of high energy hadron collisions, where detection noise in the form of reduced efficiency or limited acceptance has been applied.

The tests demonstrate that the results from the method have much smaller biases compared to the simple method described in the beginning. Furthermore the method does not exhibit any dependence on event-generator which is also in contrast to the simple method.

The full derivation and supporting verifications can be found in [6].

## Acknowledgements

We would like to express our gratitude to the Villum foundation, the Danish National Research Foundation (DNRF) and the Danish Natural Science Research Council (FNU) for their financial support for this research.

## References

- [1] T. Sjostrand, M. van Zijl, Phys.Rev. **D36**, 2019 (1987)
- [2] R.C. Hwa, C. Yang (2007), **0705.3073**
- [3] K. Wraight, P. Skands, Eur.Phys.J. **C71**, 1628 (2011), **1101.5215**
- [4] T. Sjostrand, S. Mrenna, P.Z. Skands, JHEP **0605**, 026 (2006), **hep-ph/0603175**
- [5] P.Z. Skands (2009), **0905.3418**
- [6] K. Gulbrandsen, C. Soegaard (2014), **1408.3391**