H→ e+e− at CEPC: ISR effect with MadGraph

Tanyh

2018.11.30

CEPC basic information:

- Center-of-mass energy: 240-250GeV
- Integrated luminosity: $5ab^{-1}$. 10 years:10^6 Higgs
- Simulation: GEANT4-based full simulation samples
- The signal process is generated by MadGraph, with Initial State Radiation (ISR) implemented

Result:

- The upper limit of B(H→e+e-) could reach 0.024% at 95% confidence level.

Motivation:

- Many new physics models predict the Higgs couplings deviate from the SM at the percent level.
- With this consideration, a Higgs factory at e+e− collider with high luminosity is best suited for this goal, due to its clean environment and relative lower cost.

ZH,H->e+e-:

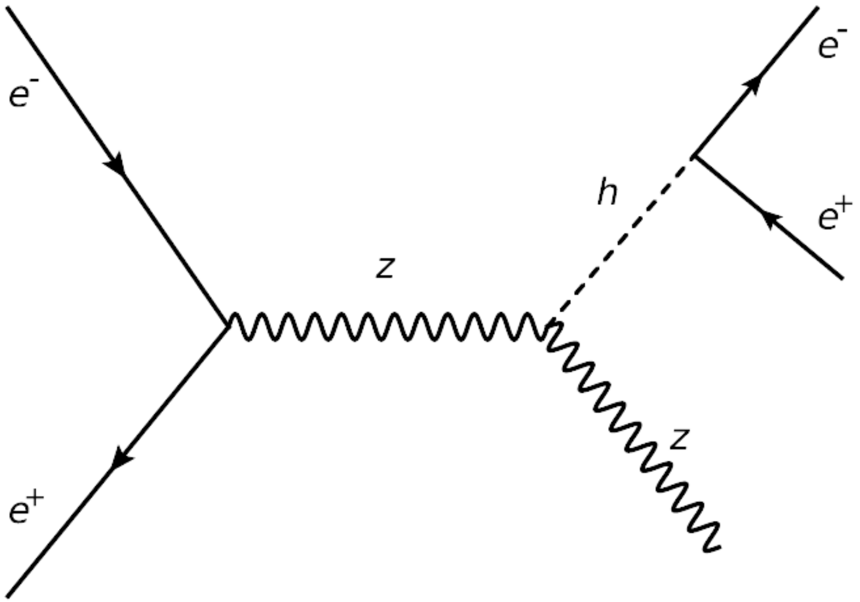


Figure 1. Example Feynman diagram.

- The SM prediction for the branching fraction B(H→e+e-) is as tiny as approximately 5×10^{-9}.

- In new physics scenario, it can be enhanced significantly.

Question from Yuzhen:
In Figure2, why does the curves decrease suddenly at about 250GeV?

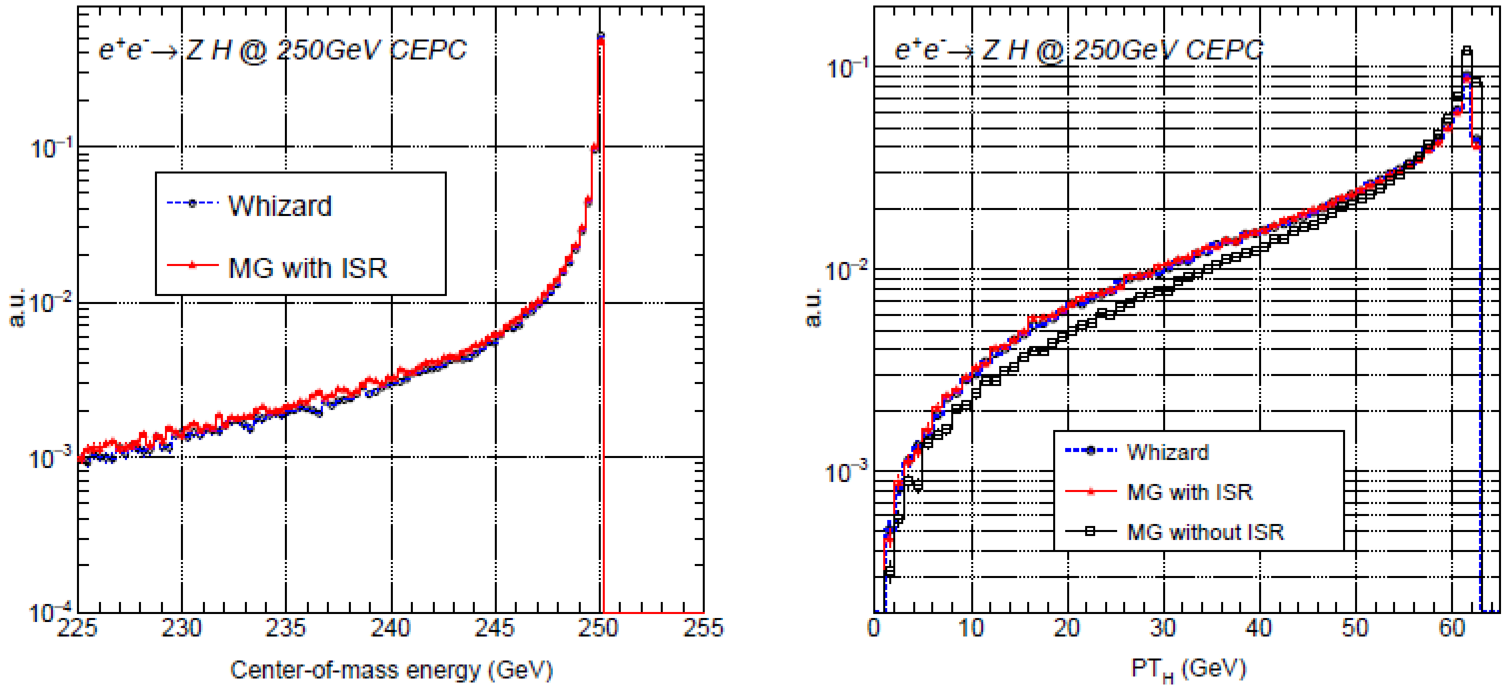
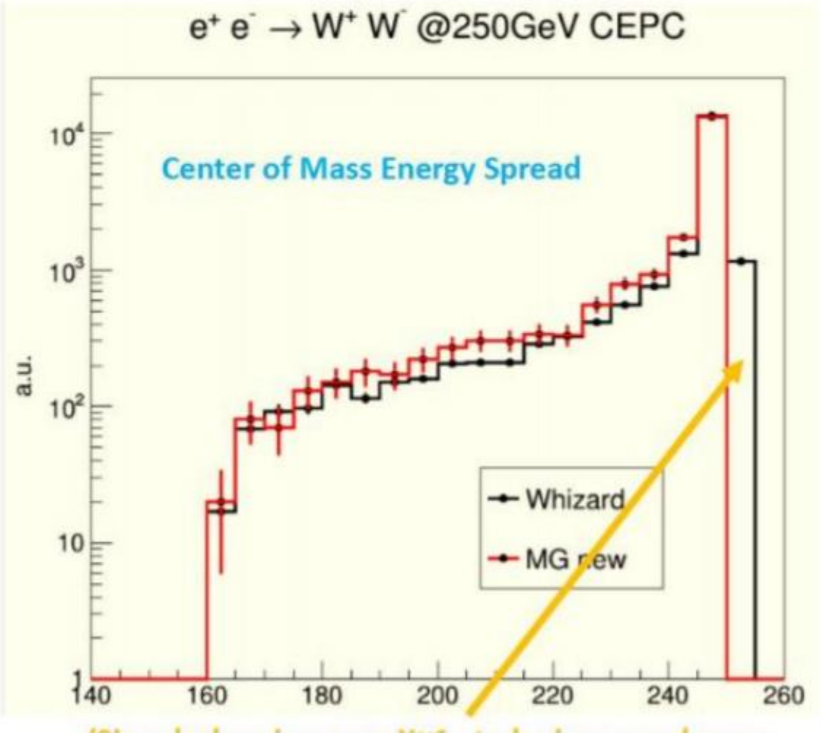


Figure 2. Comparisons plots on center-of-mass energy and Higgs transverse momentum, between WHIZARD and MADGRAPH with or without ISR effect included, for the process $e^+e^- \to ZH$.

Answer:

I don't know. But maybe the author only simulate 160Gev to 250Gev.



Question from Ryuta:

The ISR implementation in MadGraph is the title in section 2, but, could you briefly introduce the "Whizard" , "MadGraph" and their relationship ?

Answer:

MadGraph5_aMC@NLO is a framework that aims at providing all the elements necessary for SM and BSM phenomenology, such as the computations of cross sections, the generation of hard events and their matching with event generators, and the use of a variety of tools relevant to event manipulation and analysis.-**High precision simulation**

WHIZARD is a program system designed for the efficient calculation of multi-particle scattering cross sections and simulated event samples.

Question from Amit:

How will the MADGRAPH change the effect of ISR in this analysis?

Answer:

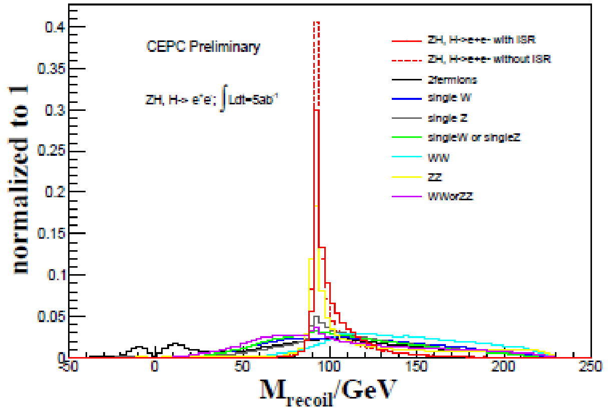
MadGraph is a simulation method, and it can simulate e+e- ->ZH with ISR and without ISR respectively. I don’t know how to use MadGraph.

Recoil mass:

The two electrons from Higgs decay can be easily identified and their momentum can be precisely measured in the detector. The Higgsstrahlung events can then be reconstructed with the recoil mass method:

$$m_{recoil}^2 = s + m_H^2 - 2 \cdot E_H \cdot \sqrt{s} \ , \tag{1.1}$$

where $\sqrt{s}$ is the center of mass energy, m_H and E_H are the mass and energy of the Higgs boson reconstructed by the two lepton four momentum. Therefore, the ZH $(H \to e^+e^-)$ events form a peak in the M_{recoil} distribution at the Z boson mass. With the recoil mass method, the ZH events are selected without using the decay information of the Z boson.



Signal and background:

For the signal process, $e^+e^- \to ZH$ with $H \to e^+e^-$, 50K events are generated by MadGraph V2.3.3 with ISR effect included, with Higgs mass set to be 125 GeV. For the backgrounds, Whizard V2.2.8 [14], are exploited as the event generator. All these samples are produced at the center-of-mass energy of 250 GeV.

The major SM backgrounds, including all the 2-fermion processes($e^+e^- \to f\bar{f}$, where $f\bar{f}$ refers to all lepton and quark pairs except $t\bar{t}$) and 4-fermion processes(ZZ, WW, ZZ or WW, single Z, single W). The initial states radiation (ISR) and all possible interference

Cut condition:

- At least one pair of electrons with opposite charge is required
- The pair with invariant mass Me+e− closer to Higgs mass is selected in case of multi-combinations, and required then to satisfy 120 < Me+e− < 130GeV.
- The recoil mass Mrecoil of e+e− is required to be greater than 90 GeV and less than 93 GeV, to be consistent with the Z-boson hypothesis.
- To suppress 2-fermions background, it is required that the difference between the two electrons' azimuth angles should satisfy $\Delta\phi < 166^\circ$.

total
$N_{e^+} \geq 1,\ N_{e^-} \geq 1$
$120\ \text{GeV} < M_{e^+e^-} < 130\ \text{GeV}$
$90\ \text{GeV} < M_{recoil} < 93\ \text{GeV}$
$46\ \text{GeV} < p_{\text{T}e^+e^-} < 63\ \text{GeV}$
$-42\ \text{GeV} < p_{\text{Z}e^+e^-} < 41\ \text{GeV}$
$\Delta\phi < 166^\circ$
$cos_{e^+} \geq -0.07,\ cos_{e^-} \leq 0.14$

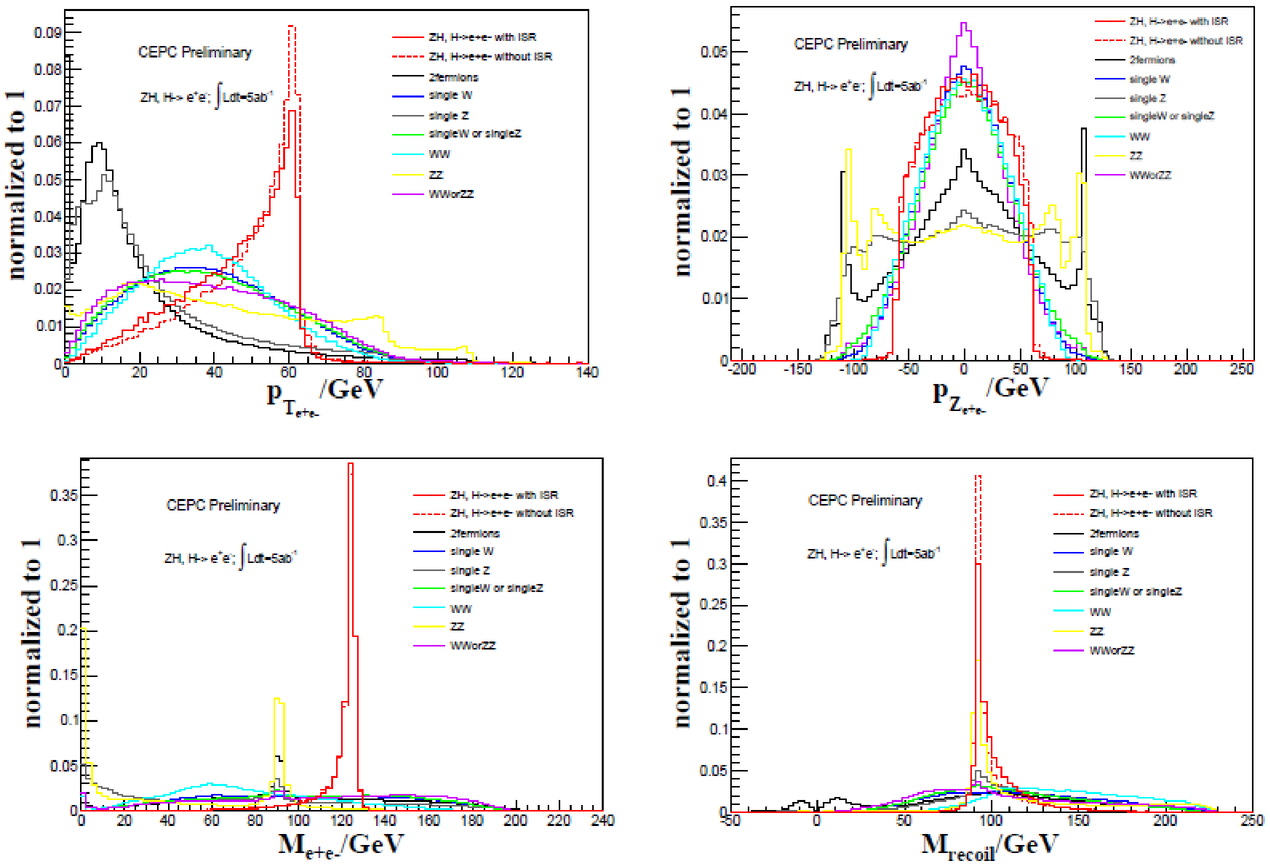


Figure 3. Distributions of $p_{\mathrm{Te^+e^-}}$, $p_{\mathrm{Ze^+e^-}}$, $M_{e^+e^-}$ and M_{recoil} for signals and backgrounds. Signals without ISR effect included are also superimposed for comparison.

Cut:

ZH,H->e+e- only this channel, Unnormalized

Category	signal	2fermions	single ZorW	single Z	single W
total	50000	418194802	1259165	7913405	17190655
$N_{e^+} \geq 1$, $N_{e^-} \geq 1$	47418	36822471	978594	3480494	2260761
120 GeV $< M_{e^+e^-} <$ 130 GeV	34463	1954192	71193	126094	151950
90 GeV $< M_{recoil} <$ 93 GeV	12362	61089	3564	6954	7255
46 GeV $< p_{\mathrm{T}e^+e^-} <$ 63 GeV	8582	6816	1863	1861	3652
-42 GeV $< p_{\mathrm{Z}e^+e^-} <$ 41 GeV	8511	6372	1783	1750	3468
$\Delta\phi < 166°$	7404	5131	1696	1651	3233
$cos_{e^+} \geq$ -0.07, $cos_{e^-} \leq$ 0.14	3564	241	86	48	161

Category	WW	ZZ	WWorZZ	total background
total	49115769	4967152	21902983	520543931
$N_{e^+} \geq 1$, $N_{e^-} \geq 1$	640839	758732	814608	45756499
120 GeV $< M_{e^+e^-} <$ 130 GeV	26731	7593	55196	2392949
90 GeV $< M_{recoil} <$ 93 GeV	1783	1464	2434	84543
46 GeV $< p_{\mathrm{T}e^+e^-} <$ 63 GeV	868	682	1297	17039
-42 GeV $< p_{\mathrm{Z}e^+e^-} <$ 41 GeV	837	647	1247	16104
$\Delta\phi > 166°$	702	566	1182	14161
$cos_{e^+} \geq$ -0.07, $cos_{e^-} \leq$ 0.14	20	178	70	804

Question from kai:

Last line of page4, 50000 signal generated and final efficiency about 7%, this will give a yield about 3500, but only about 20 in Figure 4, what's the reason?

Answer: Generate 10^6 Higgs. BR(H->e+e-)=0.024%.
The number of H->e+e- generated is 240.
The efficiency of selecting is 7%.
So the number left is 7%$\times$240=17.

Question from Suyu:

It says that 'The 95% confidence level upper limit on H->ee can then be decided to be 0.024%.' Could you explain the process in which they get the result?

Question from Xin:

In the page5, it says " the 95% confidence level upper limit on $H \rightarrow e^+e^-$ branch ratio can then be decided to be 0.024%.", it looks the 0.024% is an output value from the calculation. But why on page 6, Figure 4, in the caption, it says "assuming" the BR = 0.024%?

Calculate upper limit:

After the event selections as mentioned above, we perform a $\mu S + B$ fit (with μ as the signal strength) on CEPC simulated data which is essential purely background as the SM predicted $H \to e^+e^-$ branch ratio is too low. As shown in Fig. 4, an unbinned maximum

predicted $H \to e^+e^-$ branch ratio is too low. As shown in Fig. 4, an unbinned maximum
likelihood fit is performed on $M_{e^+e^-}$ spectrum, in the region of 120 GeV to 130 GeV. The Higgs signal shape is described by a Crystal Ball function, while the background is represented by a second order Chebychev polynomial function, whose parameters are fixed to the values extracted from the background samples. By scanning over signal strength in the $\mu S + B$ fit, one can extract the dependence of negative log likelihood on it. The 95% confidence level upper limit on $H \to e^+e^-$ branch ratio can then be decided to be 0.024%.

1.Get the distribution of μ which obey Gaussian.
2.Get the 95%CL upper limit of μ.
3.Get the 95%CL upper limit of signal branch.
4.Use the 95%CL upper limit of signal branch to fit ZH_inclusive.

This corresponds to a signal yield of around 20, while from Figure 4, the background yield under the Higgs peak is near 200, and thus by naively couting, $S/\sqrt{B} \sim 1.4$ which supports the above result from the shape analysis. Finally we mention that checks with different

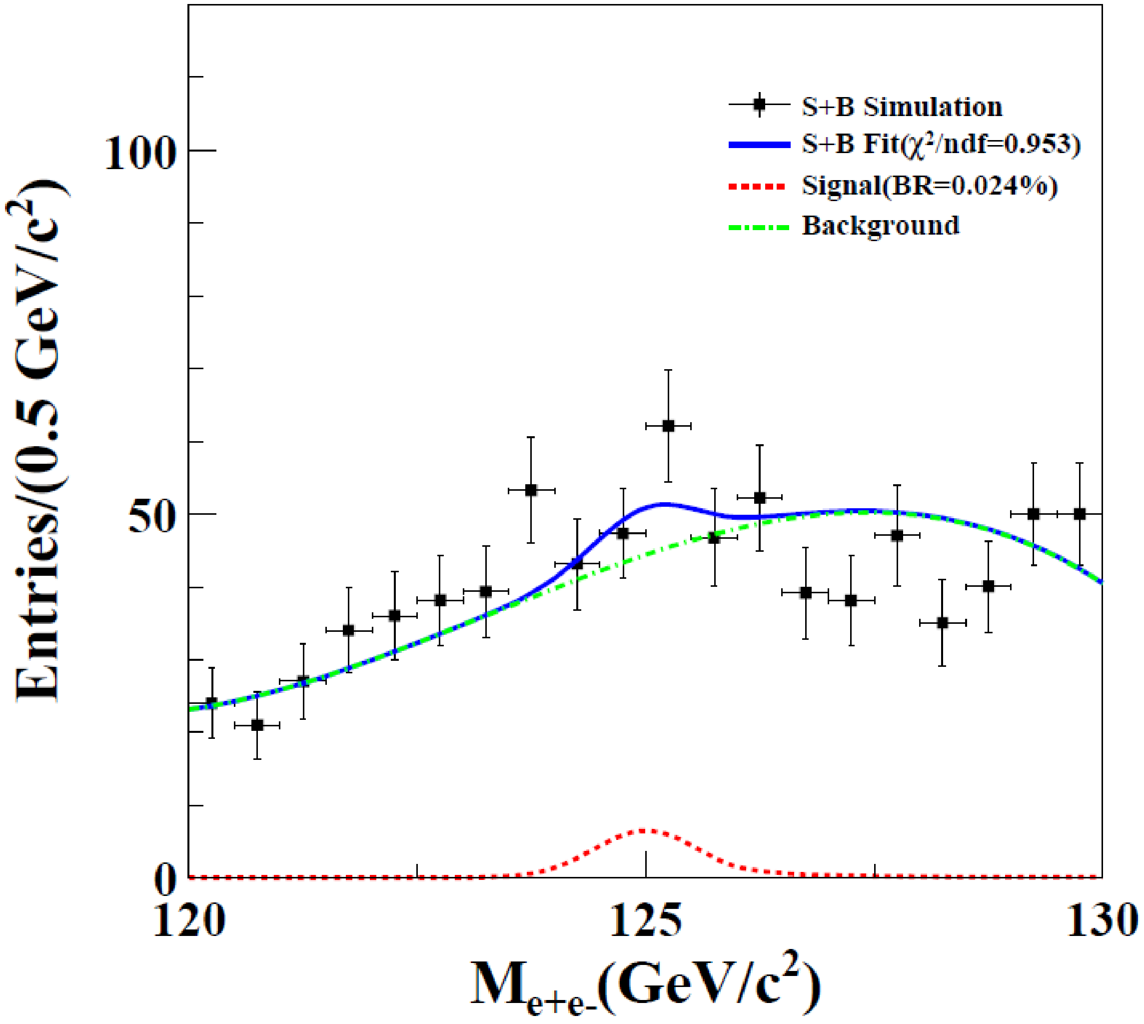


Figure 4. The invariant mass spectrum of e^+e^- in the inclusive analysis. The dots with error bars represent data from CEPC simulation. The solid (blue) line indicates the fit. The dashed (red) shows the signal (assuming $\mathcal{B}(H \to e^+e^-)$=0.024%) and the long-dashed (green) line is the background.

How to calculate upper limit:

Suppose for each event in the signal sample one measures a variable x and uses these values to construct a histogram $\mathbf{n} = (n_1, \ldots, n_N)$. The expectation value of n_i can be written

$$E[n_i] = \mu s_i + b_i \ , \tag{2}$$

$$q_0 = \begin{cases} \hat{\mu}^2/\sigma^2 & \hat{\mu} \geq 0 \ , \\ 0 & \hat{\mu} < 0 \ , \end{cases} \tag{47}$$

where $\hat{\mu}$ follows a Gaussian distribution with mean μ' and standard deviation σ. From this one can show that the pdf of q_0 has the form

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right] . \tag{48}$$

From Eq. (48) the corresponding cumulative distribution is found to be

$$F(q_0|\mu') = \Phi\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right) . \tag{50}$$

My thoughts:

95%CL: F($q_0|\mu'$)=0.95. Φ^{-1}（0.95）=1.64.

So $\breve{\mu}/\sigma$ -μ'/σ=1.65 , $\breve{\mu} > 0$

So we can get the 95%CL upper limit $\breve{\mu}$.

So we can get the 95%CL upper limit of signal

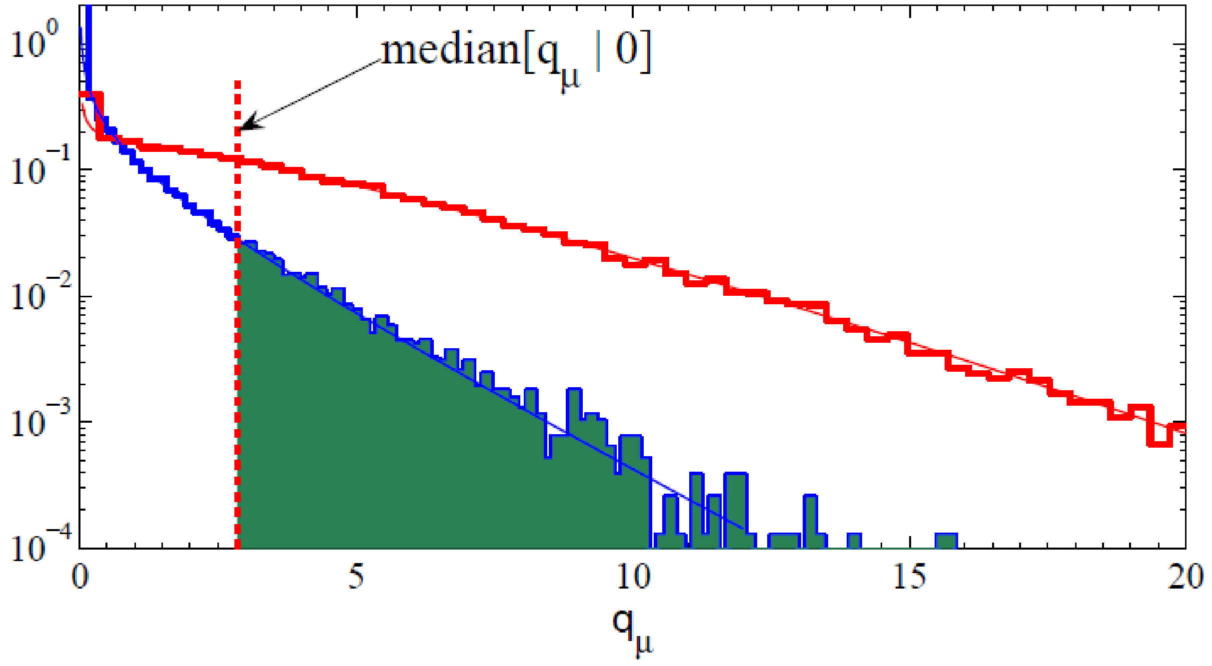


Figure 9: The distributions $f(q_\mu|0)$ (red) and $f(q_\mu|\mu)$ (blue) from both the asymptotic formulae and Monte Carlo histograms (see text).

The vertical line in Fig. 9 gives the median value of q_μ assuming a strength parameter $\mu' = 0$. The area to the right of this line under the curve of $f(q_\mu|\mu)$ gives the p-value of the hypothesized μ, as shown shaded in green. The upper limit on μ at a confidence level $\mathrm{CL} = 1 - \alpha$ is the value of μ for which the p-value is $p_\mu = \alpha$. Figure 9 shows the distributions for the value of μ that gave $p_\mu = 0.05$, corresponding to the 95% CL upper limit.

Thank you.