

BELLE 實驗之 B 介子稀有衰變分析

張寶棣、王正祥、陳凱風

國立台灣大學/物理系/高能實驗室、國立聯合技術學院

前言

位於日本國家高能中心 KEK 的 B 介子工廠 BELLE 實驗，是利用 8GeV 的電子與 3.5GeV 的正電子對撞，以產生大量的 b 夸克及反 b 夸克的共振態 $\Upsilon(4S)$ ，並藉由 $\Upsilon(4S) \rightarrow B\bar{B}$ 的事件來研究 B 介子的衰變行為。我們台大高能實驗組 (NTUHEP) 負責許多的 B 介子稀有衰變模式，如 $B \rightarrow K\pi^0$ 、 $\rho\pi$ 、 $\eta'K$ 和 $\omega\pi$ 等。而以下就以 $B^+ \rightarrow K^+\pi^0$ 模式來作一簡介。

$B^+ \rightarrow K^+\pi^0$ 模式的衰變率 (Branching Ratio) 約為 1.2×10^{-5} [1]，由 CLEO 實驗組於 1998 年發表。在正負電子對撞的環境之下，由於彈性 (elastic) 及非彈性 (inelastic) 碰撞反應的物理機制下，會有不同的物理反應結果。不同的反應機制會導致我們在探測器中偵測到不同的反應分佈。在 KEKB 對撞機的能量下，會有以下主要的反應結果：Bhabha ($e^+e^- \rightarrow e^+e^-$)、Muon-pair ($e^+e^- \rightarrow \mu^+\mu^-$)、Tau-pair ($e^+e^- \rightarrow \tau^+\tau^-$)、雙光子物理 ($e^+e^- \rightarrow e^+e^- \gamma\gamma$)、 $\Upsilon(4S) \rightarrow B\bar{B}$ ($e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$) 與非 B 介子的夸克、反夸克 ($q\bar{q}$) 事件 ($e^+e^- \rightarrow q\bar{q} \rightarrow 2 \text{ jets}$)。就 B 物理而言，經過觸發 (trigger) 篩選後，主要選取的事件為 $\Upsilon(4S) \rightarrow B\bar{B}$ ($e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$) 與非 B 介子的夸克、反夸克 ($q\bar{q}$) 事件 ($e^+e^- \rightarrow q\bar{q} \rightarrow 2 \text{ jets}$)，其截面積 (Cross-section) 比約為 1:2.8，再考慮事件重建的效率 (Reconstruction Efficiency)，使得我們在每一百萬個事件中，大約只能找到一個左右真正的 $B^+ \rightarrow K^+\pi^0$ 。因此我們需要應用多種不同的分析方法，如 Event Shape、Fisher Discriminator、類神經網路 (Neural Network)、以及 Maximum Likelihood Fit，在大量的數據當中分別背景 (background) 事粒及尋找我們所需的 B 事件。

事件模擬 (Monte Carlo) 與重建 (Reconstruction)

就模擬而言其主要有兩部分：其一為偵測器模擬 (GSIM)，主要模擬粒子在探測器的反應。另一為模擬物理事件產生 (QQ98)。以本分析而言，我們用 QQ98 產生三種不同事件： $B^+ \rightarrow K^+\pi^0$ 訊號事例、一般 B 介子衰變事件 (Generic $B\bar{B}$ Event) 與來自夸克、反夸克的背景 (Continuum Background) 的 Monte

Carlo 資料，作為測試各種分析方法的基礎。

就重建而言，在探測器中只會看到粒子衰變後生命期較長的粒子。在 BELLE 中只有下述粒子 (包括反粒子) 會被看到：光子 (photon)、電子 (electron and positron)、質子 (proton and antiproton)、渺子 (muon)、帶電 K 介子及 π 介子。我們使用 BELLE 偵測器的 Central Drift Chamber (CDC) 來得到帶電粒子的軌跡，並由其在磁場中的曲率及磁場大小而

得知其動量。並利用 BELLE 的粒子鑑別系統(PID)，Aerogel Cerenkov Counter(ACC)、Time of Flight(TOF)、dE/dX，判別帶電粒子的質量而得到其能量。藉由 PID 我們可以分別 K^+ 及 π^+ 。光子及渺子分別有偵測器來偵測它。而 π^0 介子可以從 $\pi^0 \rightarrow \gamma\gamma$ 衰變模式以兩顆光子的四維動量(4 Momenta)相加，再假設 π^0 的質量，作 Mass Constraint Fit 獲得。

分析(Analysis)

我們通常使用倒推式的方法來找尋我們所要的衰變模式。例如要找 $B^+ \rightarrow K^+ \pi^0$ ，我們先找 π^0 及 K^+ 介子再將其能量及動量相加，若此 π^0 及 K^+ 介子為從 B 介子來的，其質量將會落在 B 介子的質量範圍。另可以預見的是在眾多 π^0 及 K^+ 介子中也可能有隨

機組合的事件落在 B 介子的質量範圍。而下列一系列的方法，就是為了要儘量減少此機會。

的在此我們定義兩個變數 ΔE 與 Beam Constrained Mass(M_{bc})，作為選擇 $B^+ \rightarrow K^+ \pi^0$ 事例的依據：

$$\Delta E = E\{K, \pi^0\} - E_{Beam}$$

$$M_{bc} = \sqrt{E_{Beam}^2 - P^2\{K, \pi^0\}}$$

其中 E_{Beam} 為質心系統能量的一半，約為 5.29GeV。利用 ΔE 與 M_{bc} 這兩個變數，我們可以定義兩個區域：訊號區(Signal Region)和邊緣帶(Grand Side Band)(不包含 Signal Region)，見表一與圖一。

表一：訊號區(Signal Region)和邊緣帶(Grand Side Band)定義

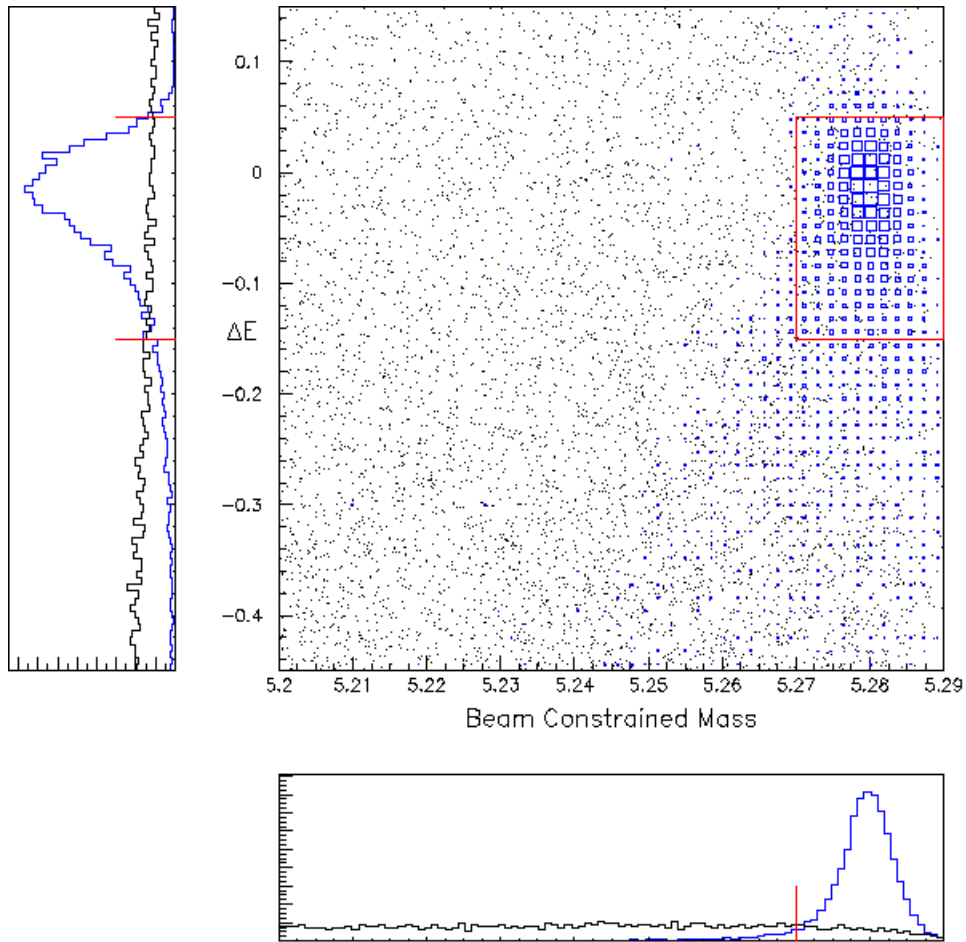
	ΔE	M_{bc}
Grand Side Band	-0.45 to 0.15	>5.2
Signal Region	-0.15 to 0.05	>5.27

如果我們只計數 Signal Region 部分的事件，可以得出約 33.6%的訊號重建效率，但是可以消除 Continuum Background 到萬分之一左右，以及幾乎全部的一般 B 介子背景(Generic $B\bar{B}$ Background)，約可以達到 0.11 的訊雜比(S/N Ratio)。此數目仍不足以讓我們宣稱找到此種衰變模式，Continuum Background 仍然太大。

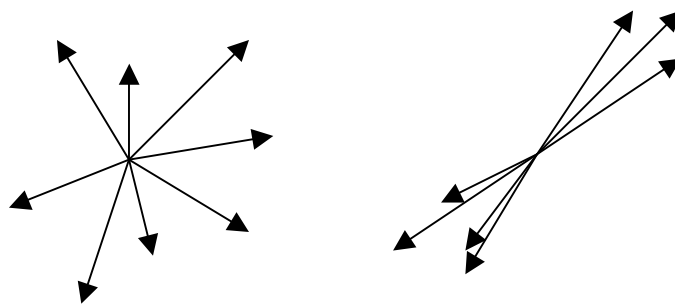
Event Shape 變數

再接下來我們就必須要用到與 B 介子與夸克、

反夸克的空間分布特性，來消除背景事件。在 B 介子的事例中，空間中粒子的來源都是從這兩顆 $B^+ B^-$ 或 $B^0 \bar{B}^0$ 而來。以 $B^+ \rightarrow K^+ \pi^0$ 為例，如果我們將從 B^+ 衰變得來的 K^+ 、和從 π^0 二次衰變成的兩顆光子抽出，而剩下來的粒子就都是從另一顆 B 介子得來(在這裡就是 B^-)。這些剩下的粒子會在質心座標系中呈現球形分布。在夸克、反夸克的事件中，粒子會形成兩束射出狀的分布(Jet-Like)。如圖二：



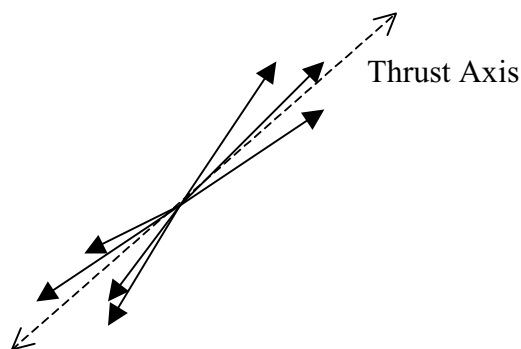
圖一：其中的 Box plot 是 Signal 的 Monte Carlo 模擬，而點圖代表著 Continuum Background 模擬，而矩形的區域即為 Signal Region。



圖二：B 介子對(左)與夸克、反夸克(右)的空間分布特性。

根據此一特性，我們可以再空間中尋找一個方向軸，使所有的粒子(不包含被訊號用去的粒子，即為

K^0 與兩顆光子)在上面的動量投影為最大值，即稱為 Thrust Axis(如圖三)。



圖三：所有的粒子在 Thrust Axis 上面的動量投影為最大值，但不包含 K^+ 與兩顆 Gamma。

根據 Thrust Axis、和 B 介子與夸克、反夸克的空間分布特性，再套用 B 介子的動力學，我們可以定出四個 Event Shape 變數：

1. $\cos\theta_{thrust}$ ：Thrust Axis 和訊號部分 π^0 夾角的餘弦值。

2. Sphericity(S_{\perp})： $S_{\perp} = \frac{\sum P_t}{\sum P}$ ， P_t 是垂直

Thrust Axis 的動量分量，但不包含與 Thrust Axis 夾角小於 45 度的粒子。

3. R_2 ：第二階 Fox-Wolfram moment^[2]，定義：

$$H_l = \frac{|p_i| \cdot |p_j|}{E_{total}} P_l(\cos\theta_{ij})、R_l = H_l / H_0。$$

其中 P_l 是 Legendre Polynomials。

4. $\cos\theta_B$ ：B 介子和 Z 軸(約為 Beam 的方向)夾角的 Cosine 值。

以上的四個變數的分布情形可以從 Monte Carlo 模擬得到(圖四)。運用 ΔE 、 M_{bc} ，以及四個 Event Shape 變數，我們可以得到一個 Slice-Cut 的結果(即為設定切割範圍，直接去除背景)(表二)。

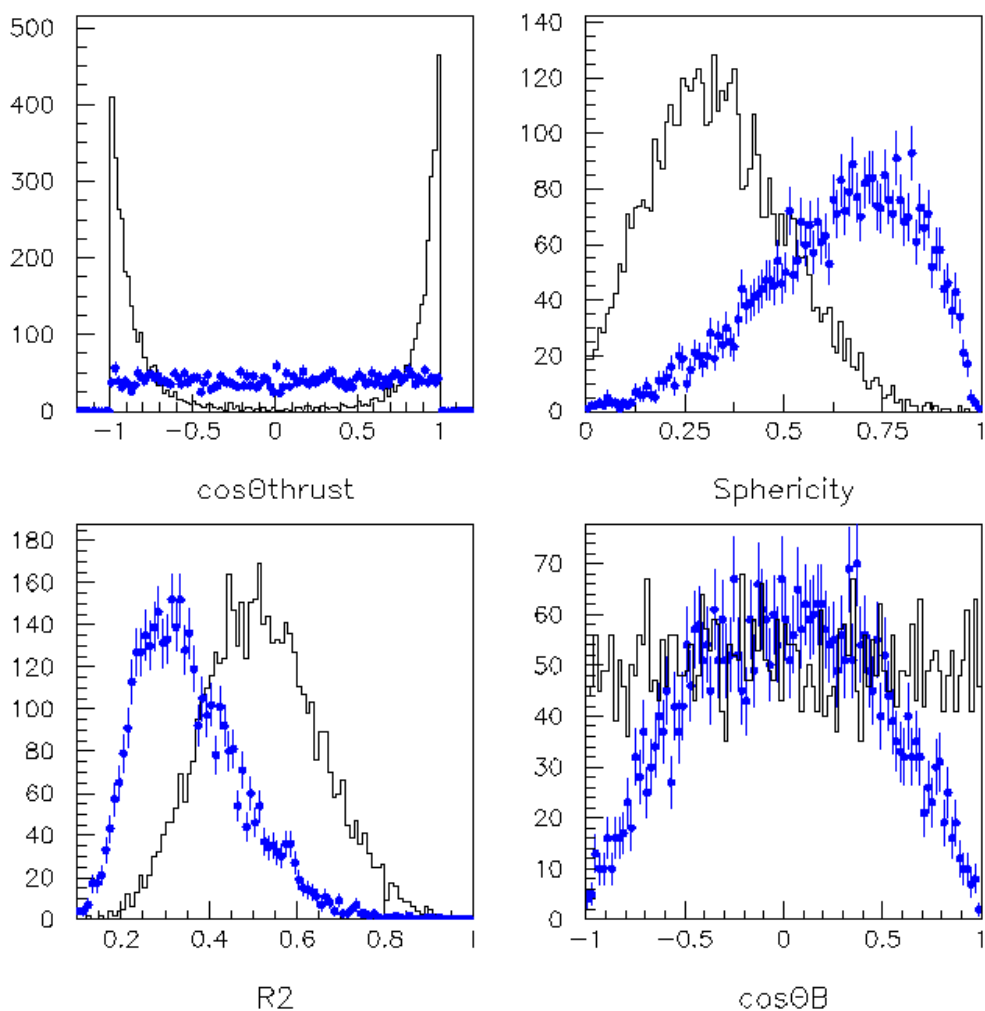
表二： $B^+ \rightarrow K^+ \pi^0$ 的 Slice-Cut 結果

Cuts	$K^+ \pi^0$ signal (Sig. Eff.)	Pre-selected cont.	$B\bar{B}$ generic	April 30 th cont.	Sept. 3 rd cont.
Data samples	20,000	29.18M	1.0M	2.03M	20M
$K^+ \pi^0$ Side Band & Signal Region	11,172(56%)	27,901	1	2078	19582
PID($K\pi$)	7987(40%)	7460	0	553	5443
Signal region	6730(32%)	384	0	32	268
$\cos\theta_{thrust} < 0.8$	5383(27%)	76	0	6	48

Sphericity > 0.6	4103 (20.5%)	23	0	2	19
R2 < 0.4	3708 (18.5%)	20	0	1	15
$\cos\theta_B < 0.8$	3494 (17.5%)	12	0	1	10

套用所有的變數後，其訊雜比：

$$S/N \cong \frac{1 \cdot Br(\approx 1.2 \times 10^{-5}) \cdot 17.5\%}{2.8 \cdot 12 / 29.18M} \cong 1.82$$



圖四： $\cos\theta_{thrust}$ 、Sphericity(S_{\perp})、R2、 $\cos\theta_B$ 的分布情形，其中點圖是模擬的 $B^+ \rightarrow K^+ \pi^0$ 訊號事例，實線則為 Continuum Background。

最佳化方法

接下來我們討論一些演算法，和一個新式的 Event Shape 變數：Super-Fox/Wolfram，可以用來最佳化我們的結果。

I.) Fisher Discriminator

Fisher Discriminator 最主要的想法是一種一維線性投影。經由矩陣運算，可以得到一些變數的線性組合，使得最後的一維分布上，訊號和背景分布之間的距離最遠。

定義： $F = \vec{\alpha} \cdot \vec{x}$ 。其中， \vec{x} 為輸入參數的向量

(Row Vector)， $\vec{\alpha}$ 為這些參數的係數。最佳化 $\vec{\alpha}$

的計算方法如下：

1. 準備輸入參數的事件(訊號和背景)和他們

的平均向量： \vec{M}_s 、 \vec{M}_b 。

2. 計算誤差矩陣(Mean-Square Error)

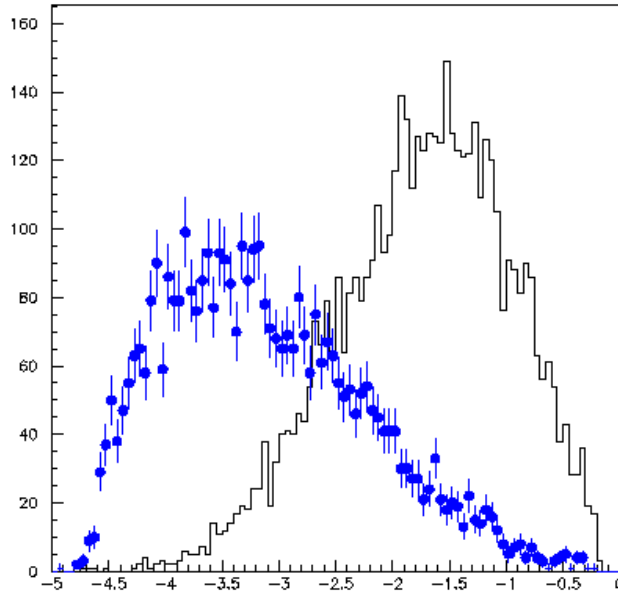
$$U_s = \left\langle \left(\vec{x}_s - \vec{M}_s \right)^2 \right\rangle$$

$$U_b = \left\langle \left(\vec{x}_b - \vec{M}_b \right)^2 \right\rangle$$

3. 於是得到最佳化的 $\vec{\alpha}$ ：

$$\vec{\alpha} = \left(\vec{U}_b + \vec{U}_s \right)^{-1} \cdot \left(\vec{M}_b - \vec{M}_s \right)$$

我們利用 $\cos\theta_{thrust}$ 、Sphericity(S_{\perp})、 R_2 當作參數，將一部份的 Monte Carlo 資料用於計算 $\vec{\alpha}$ ，再將未用過的資料輸入作為測試，Fisher 的輸出分布為圖五。

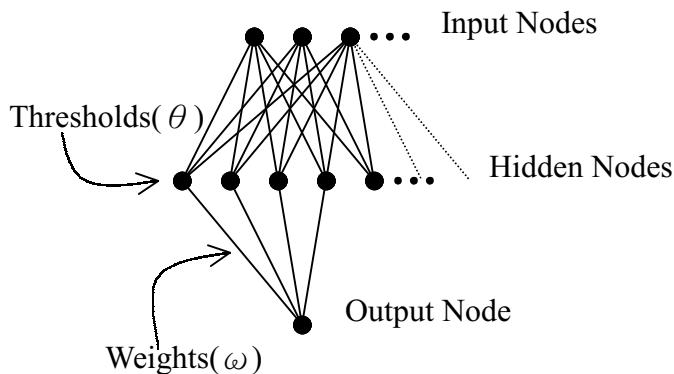


圖五：Fisher Discriminator 的輸出結果。參數為 $\cos\theta_{thrust}$ 、Sphericity(S_{\perp})以及 R_2 。其中點圖是模擬的 $B^+ \rightarrow K^+ \pi^0$ 訊號事例，實線則為 Continuum Background。

II) 類神經網路 (Artificial Neural Network, ANN)

類神經網路(ANN)是一個時常用來作為圖像辨識 (Pattern Recognition, PR)與解決非線性的問題的工具。而在這邊我們使用的是一種 Feed-Forward 型式的 Multilayer Perception Networks(MLP)，使用的軟體是由 C. Peterson 與 T. Pognvaldsson 等所發展的 JETNET 3.4^[3]。

神經網路最直接的想法可以看做是一種尋找尋找最小值的問題。藉由定義誤差，以及一些特殊的演算法，來修正神經網路中的參數，即一些 Weights 和 Thresholds。這一個動作稱為訓練 (Training)，經過反覆的訓練之後，使得輸入參數與輸出結果有正確的關聯性。如圖六，我們這裡所使用的網路結構是採用三層的節點(一層隱藏)，並使用 Manhattan 學習規則。其中使用的 Activation function 是： $g(x) = (1 + e^{-2x})^{-1}$



圖六：神經網路的架構，有輸入節點(Input Nodes)，隱藏節點(Hidden Nodes)，與一個輸出節點(Output Node)。而每一條連結線上都有 Weight，而每一個節點上都有 Threshold 與 Activation function。

如果將這一個網路寫成數學式，定義輸出 $F(\vec{x})$ ：

$$F(\vec{x}) = g\left\{\frac{1}{T} \sum \omega_{i,j} g\left[\frac{1}{T} \sum \omega_{i,k} g(x_k) + \theta_j\right] + \theta_i\right\}$$

其中， \vec{x} 為輸入向量， x_k 元素對應第 k 個輸入節點。 $\omega_{i,j}$ 代表第 I 層的第 I 個節點與第 J 層的第 j 個節點的連結 Weight， θ_i 代表第 I 層的第 I 個節點上的 Threshold。其中 T 為學習溫度 (Learning Temperature)， $g(x)$ 就是前面所寫的

Activation function。定義方均誤差 (Mean-Square Error)：

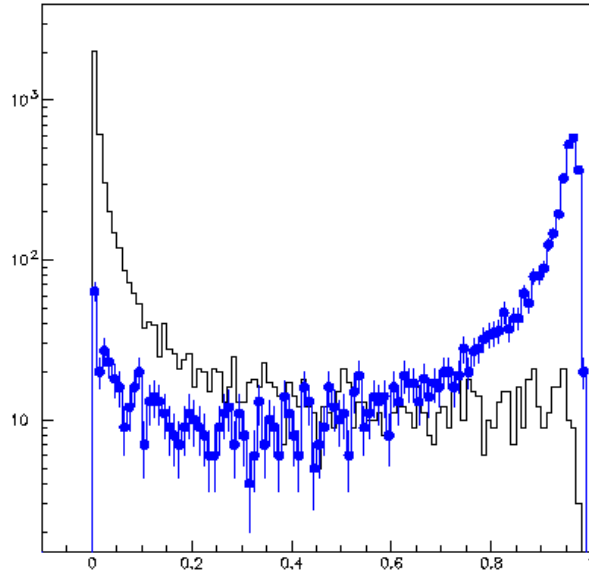
$$E = \frac{1}{2N_p} \sum \sum (o_n^{(p)} - t_n^{(p)})^2$$

$o_n^{(p)}$ 與 $t_n^{(p)}$ 分別代表實際輸出與理想值，為第 n 個 Pattern，以及第 p 個輸出節點，在我們這裡 p 只等於 1。所有的 Weights 與 Thresholds 必須由最小化誤差的動作得到，也就是學習：

$$\Delta\omega = -\eta \cdot \text{sgn}\left[\frac{\partial E}{\partial \omega}\right]$$

這就是 Manhattan Updating Rule，其中 η 是學習率(Learning Rate)。將我們一部份的 Monte

Carlo 資料用於訓練，再將未用過的資料輸入作為測試，於是我們可以獲得如圖七的分布：



圖七：神經網路的輸出結果，共使用三層的節點(6-10-1)。參數為 ΔE 、 M_{bc} ，以及四個 Event Shape 變數 $\cos\theta_{thrust}$ 、Sphericity (S_{\perp})、R2 與 $\cos\theta_B$ 。其中點圖是模擬的訊號事例，實線則為背景。

III) Super Fox-Wolfram

Super Fox-Wolfram(SFW)是一種新式的 Event Shape 變數，是一個 Fox/Wolfram 數列的變形。

原本的 Fox/Wolfram 數列：

$$H_l = \frac{\sum |p_i| \cdot |p_j|}{E_{total}} \cdot P_l(\cos\theta_{i,j})$$

$$R_l = \frac{H_l}{H_0}$$

其中， P_l 為 Legendre Polynomials。是將所有粒子的動量 p_i 與 p_j 與他們的夾角 $\theta_{i,j}$ ，作計算獲得。但是在 Super Fox-Wolfram 中，模仿其他變數的作法，將被拿來作訊號 B 介子的粒子與剩下

的粒子分開。於是這一個 R_l 中的 Σ 可以重新寫成：

$$R_l = R_l^{SS} + R_l^{SO} + R_l^{OO}$$

此時 R_l^{SS} 為訊號 B 介子的粒子與自己的計算，

R_l^{SO} 為訊號 B 介子的粒子與剩下的粒子計算，

R_l^{OO} 為剩下的粒子與自己計算。也就是將原來的

R_l 分成三個部分。因為 R_l^{SS} 與其他變數的相關性

太高，所以將他去除。將剩下來的 R_l^{SO} 與 R_l^{OO} ，

取前四階的 R-Series(即 $l=1$ to 4)，共八個變

數作線性組合。於是定義 Super Fox-Wolfram :

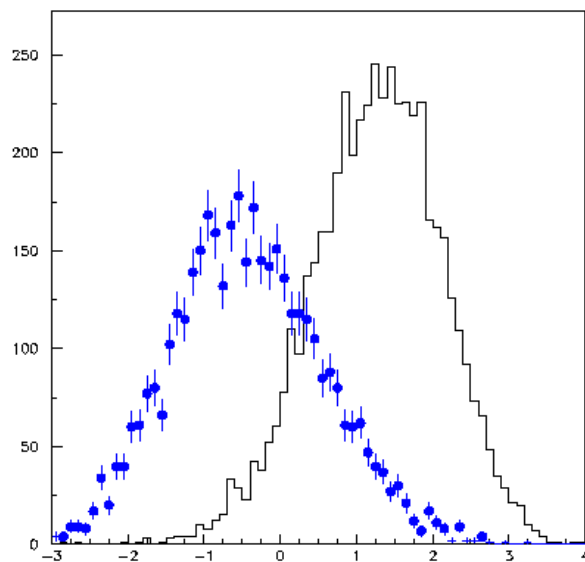
$$SFW = \vec{\alpha} \cdot \overrightarrow{R^{so}} + \vec{\beta} \cdot \overrightarrow{R^{oo}}$$

係數 \vec{a} 與 $\vec{\beta}$ 可以簡單的利用 Fisher

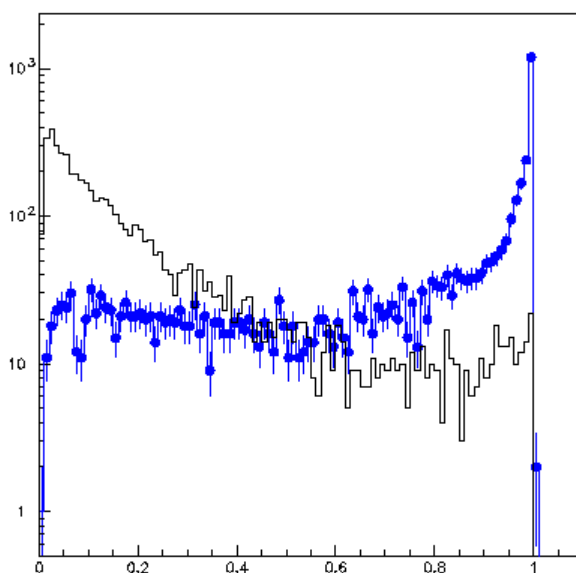
Discriminator 求得最佳化的解。或是利用

Neural Network 作非線性的組合。(見圖八與圖

九)



圖八：Super Fox-Wolfram 的輸出，是兩個 Bifurcated Gaussia 分布。其中點圖是模擬的訊號事例，實線則為背景。



圖九：Super Fox-Wolfram 的 Neural Network 輸出(nSFW)。其中點圖是模擬的訊號事例，實線則為背景。

虛擬實驗(Pseudo Experiment)

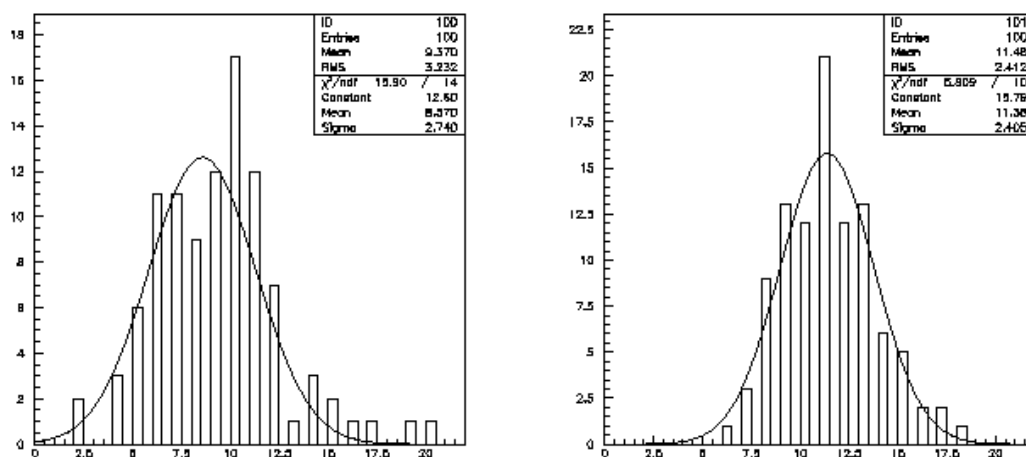
前面我們提到了許多種類的演算法，可以確實的得到不同的輸出分布，已經可以使用如前面的 Slice-Cut 一般，立刻給一個經由 Cut 的結果。但是這裡有另一種常使用的方法，可以幫助我們設定切割點，並同時預先估計真實對撞的結果，就是作一個虛擬實驗(Pseudo Experiment)。其步驟如下：

1. 準備 Monte Carlo，包含訊號與背景。然後將他們分割成兩個部分，好做訓練和測試。
2. 訓練神經網路和 Fisher Discriminator，依據“訓練樣本”。
3. 假設衰變率(Branch Ratio)約為 1.2×10^{-5} ，然後重複丟骰子，選取 $B \rightarrow K \pi^0$ 事件與

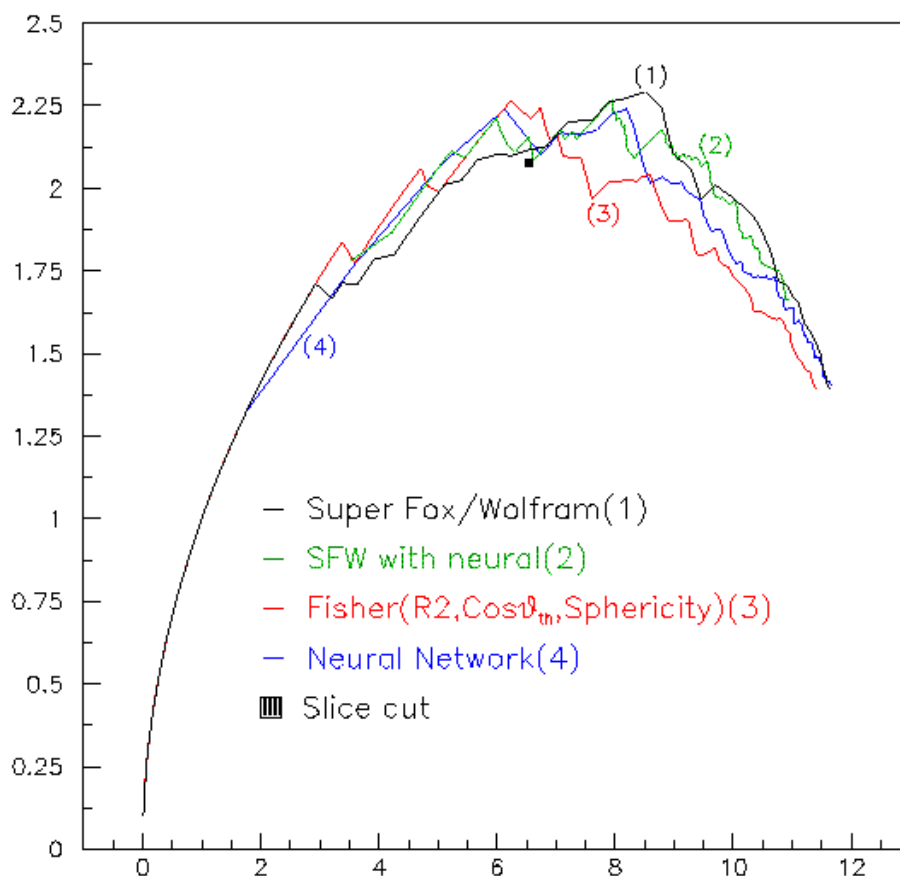
Continuum 事件的次數，得到相對應於 3fb^{-1} 的數據量(就像真正的物理實驗)。

4. 從“測試樣本”中抽出相對應數量的事例，再將他們丟進 Fisher 和神經網路等等。因為有事件重建的效率，所以我們會得到少於抽籤的樣本。
5. 於是我們得到相對應於 3fb^{-1} 的數據量的結果(註：有可能一個 $K \pi^0$ 訊號都沒有，也可能很多)。

重複以上步驟，可以得到如圖十的分布，然後套上許多不同的選擇條件與方法，最後最後可以計算出一個平均結果(圖十一)。



圖十：一個 Pseudo Experiment 的範例。左圖為經過 Neural Network 篩選的 $K \pi^0$ 訊號事例數，右為 Continuum Background 的事件數。總共有 100 次虛擬實驗。



圖十一：縱軸為 100 次 Pseudo Experiment 平均 S/\sqrt{N} 值，橫軸為訊號的事件數量。

Maximum Likelihood Fit 方法

以上我們介紹的通稱為 Slice Cut，其主要精神為找一些變數並研究其在訊號事件及背景事件，在對其變數直接取值作篩選條件。此方法的好處為簡單但缺點為效率(efficiency)不高。不同於 Slice Cut 方法，Maximum Likelihood Fit 可以對所欲研究事粒作最有效的篩選。不會如 Slice-Cut 般損失太多事例。首先定義 Likelihood L ：

$$L = P_1 \cdot P_2 \cdot P_3 \cdots P_n$$

$$P = f \cdot P_s + (1 - f) \cdot P_b$$

其中， P 為某變數在某一個事例的機率， P_s 與 P_b 分

別是此變數在訊號與背景的 Probability Density Function(PDF)，而 f 為訊號事件在全部事件中的百分比率。藉由求 L 的最大值，而解出 f (即訊號的事例量)，就稱作 Maximum Likelihood Fit。在實作上我們通常不直接處理 L ，而改考慮：

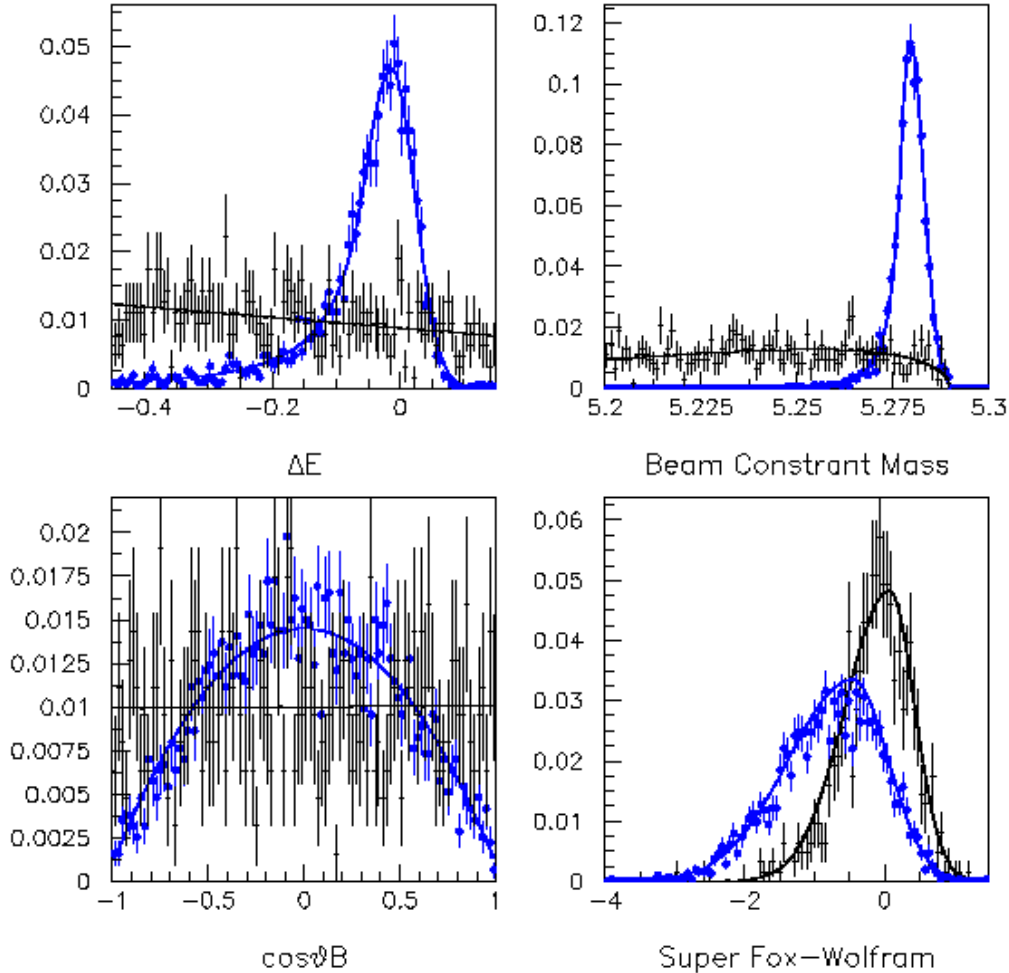
$$L' = -2 \ln(L) = \{\ln P_1 + \ln P_2 + \ln P_3 \cdots\}$$

只要計算 L' 的最小值即可。我們選取以下四個變數作為我們的 Maximum Likelihood Fit 中的變數，並選取合適的 PDF(圖十二)：

- ΔE , Energy Difference
- M_{bc} , Beam Constraint Mass

● $\cos\theta_B$, B Flight Direction

● SFW, Super Fox-Wolfram



圖十二： ΔE 、 M_{bc} 、 $\cos\theta_B$ 與 SFW 的 Probability Density Function(PDF)分布。

ΔE 、 M_{bc} 的 Signal 部分所用的 PDF 是 Crystal Ball

line :

$Cb(x) =$

$$\left(\frac{n}{a}\right)^n \cdot \exp\left(-\frac{1}{2}a^2\right) / \left(\frac{Mean-x}{\sigma} + \frac{n}{a} - a\right)^n$$

with $x > Mean - a \cdot \sigma$

$$Cb(x) = \exp\left\{-\frac{1}{2}\left(\frac{Mean-x}{\sigma}\right)^2\right\}$$

with $x < Mean - a \cdot \sigma$

SFW 的 Signal 與 Continuum Background 的 PDF 都是 Bifurcated Gaussian(左右 Sigma 不同的高斯分布)， M_{bc} 的背景形狀是 ARGUS background function :

$$A(x) = x \cdot \sqrt{1 - \left(\frac{x}{E_{Beam}}\right)^2} \cdot \exp\left\{-fact \cdot \left[1 - \left(\frac{x}{E_{Beam}}\right)^2\right]\right\}$$

但是因為背景事件在 Fit 的區域(整個 Grand Side Band 與 Signal Region)中依然非常多，所以我們仍

然需要一個 Cut，好降低背景事件的數量。否則很容易會使訊號被背景淹沒。表四中列著 100 次 Pseudo Experiment 的平均結果，假設 $3fb^{-1}$ 的數據量與 1.2×10^{-5} 的衰變率，包含兩種不同的 Cuts，極值與誤差計算的 Package 是 MINUIT。

表四：Maximum Likelihood Fit 結果

Cut	Signal Events	Signal Efficiency	Background Events
$n_{SFV} > 0.5$	$13.6^{+5.2}_{-4.8}$	37.8%	$265^{+4.8}_{-5.2}$
$ \cos\theta_{thrust} < 0.8$	$9.6^{+5.2}_{-4.3}$	26.7%	$453.0^{+4.3}_{-5.2}$

結論

以上為一通常物理分析的實例，經使用不同的方法我們可以將背景事粒減低到可以忍受的範圍。而且一如預期 Maximum Likelihood 比 Slice Cut 提高了近兩倍的效率。我們預期在 BELLE 一年內應該可以找到此一衰變模式。

參考文獻

- [1] First observation of the decay $B^\pm \rightarrow K^\pm \pi^0$, CLEO Collaboration, ICHEP98 858, CLEO CONF 98-20
- [2] G. Fox. and S. Wolfram, Phys. Rev. Letter 41 1581(1978)
- [3] C. Peterson and T. Rognvaldsson, JETNET 3.0 manual. CERN-TH. 7135/94(1993)

歡迎刊登廣告

「物理雙月刊」是一份報導物理界動態發展之刊物，其內容深入淺出，涵蓋物理新知、物理專文、人物專訪、物理消息、研討會消息等專欄，為台灣物理界人士所熟