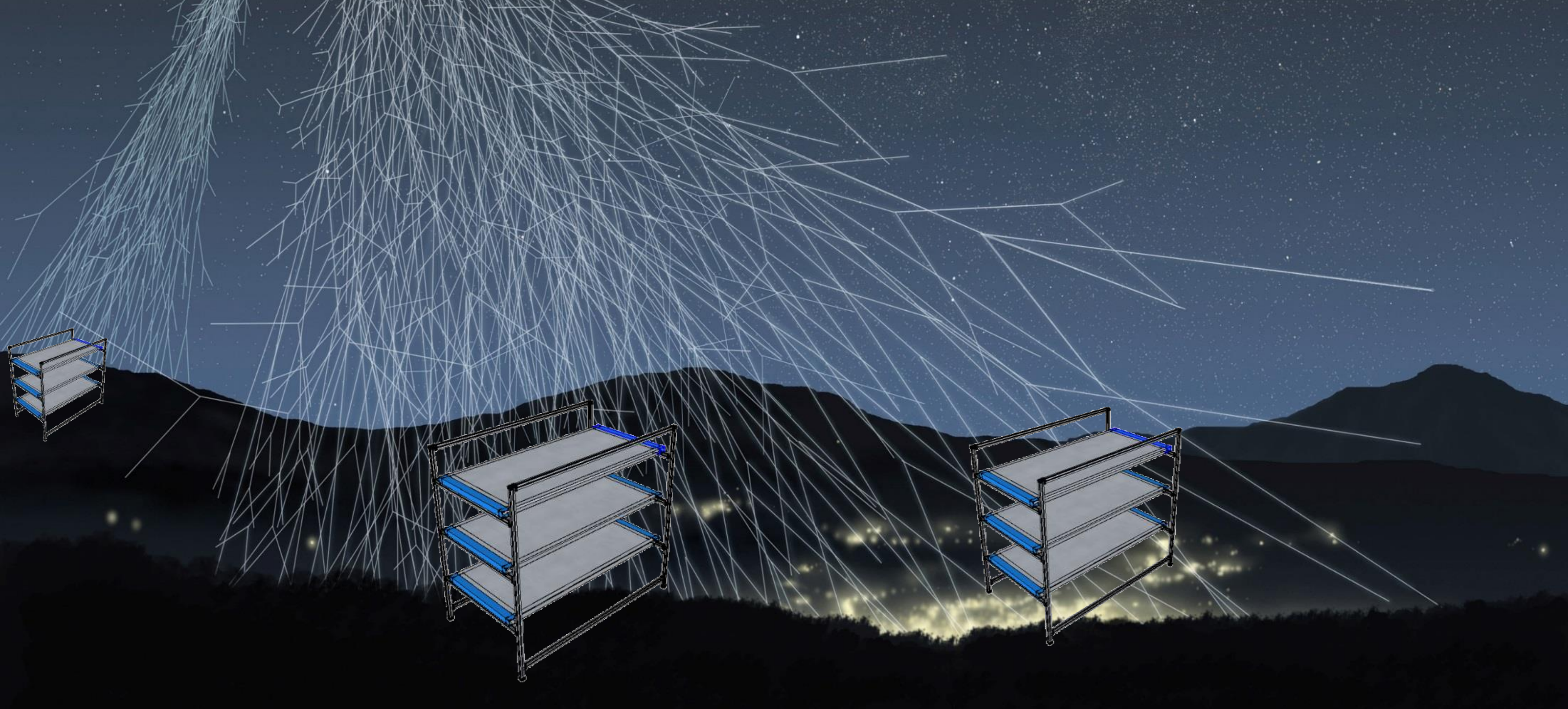
Osservazione di eventi rari e valutazione della significatività statistica

Edoardo Bossini

Erice, 28-30 Maggio 2017

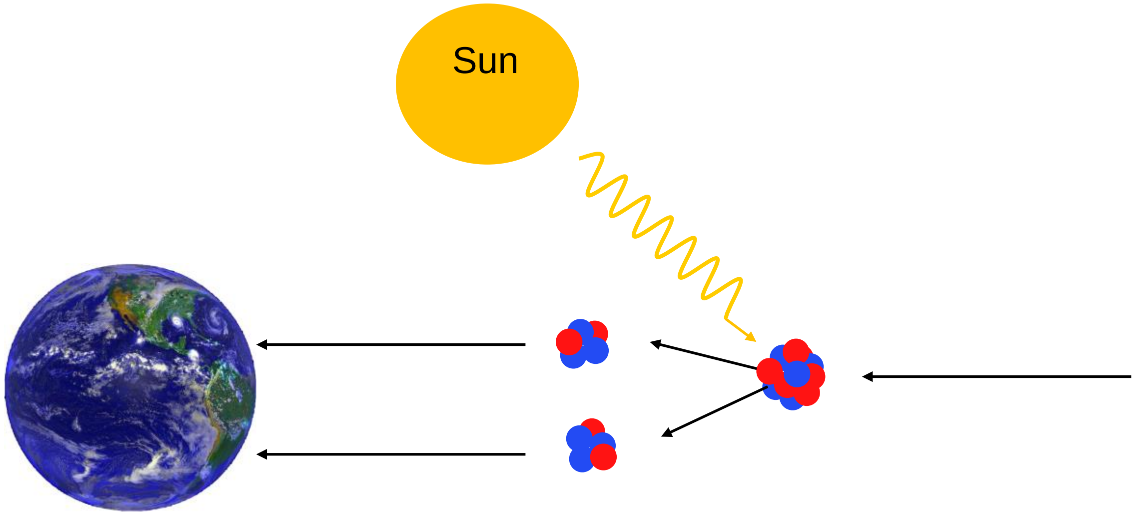


Long Distance Correlation (LDC)

In questa masterclass andremo ad analizzare le correlazioni temporali tra sciami a lunghe distanze.
In altre parole, possono due sciami distinti osservati sulla terra essere in relazione tra di loro?

Possibili meccanismi fisici in grado di giustificare tale correlazione sono stati proposti

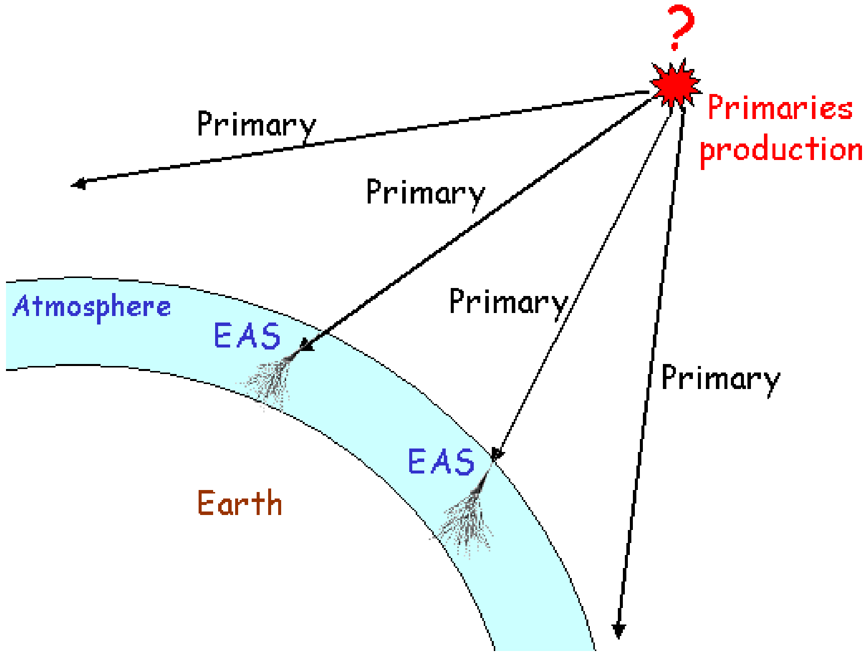
Nel modello Gerasimova-Zatsepin (GZ) la correlazione è spiegata ipotizzando un'interazione tra un nucleo e un fotone generato dal sole.



Long Distance Correlation (LDC)

Altri meccanismi sono possibili:

- Interazione con il mezzo interstellare (ricordate che lo lo spazio non è vuoto!)
- Emissione sincrona da una sorgente



Tutti questi eventi sono caratterizzati da uno o più sciami che si sviluppano ~simultaneamente sulla terra

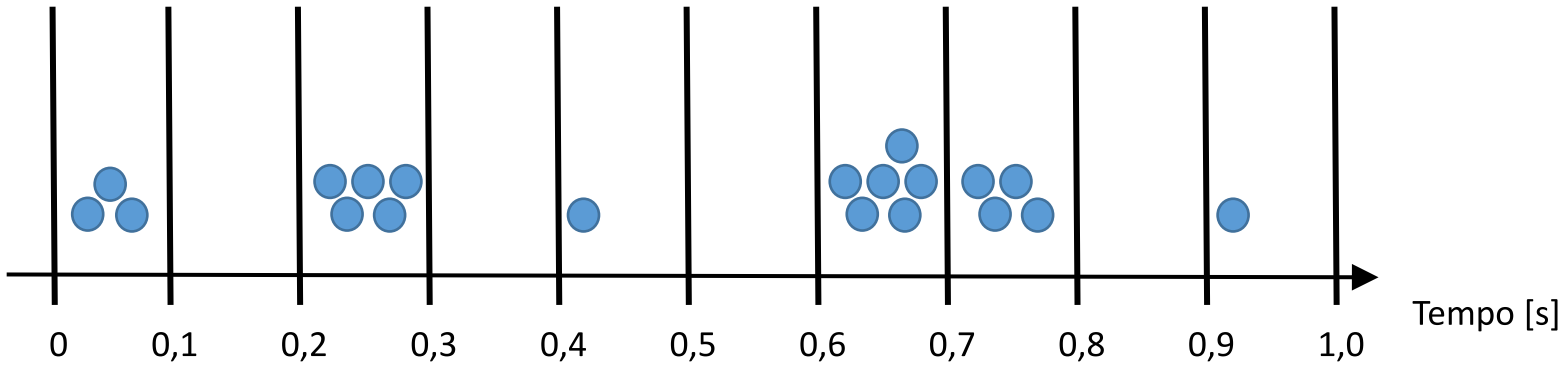
Ma quanti eventi LDC ci sono?

Gli eventi sono molto rari e si stimano compresi tra 1-0,001/anno/km^2!!!

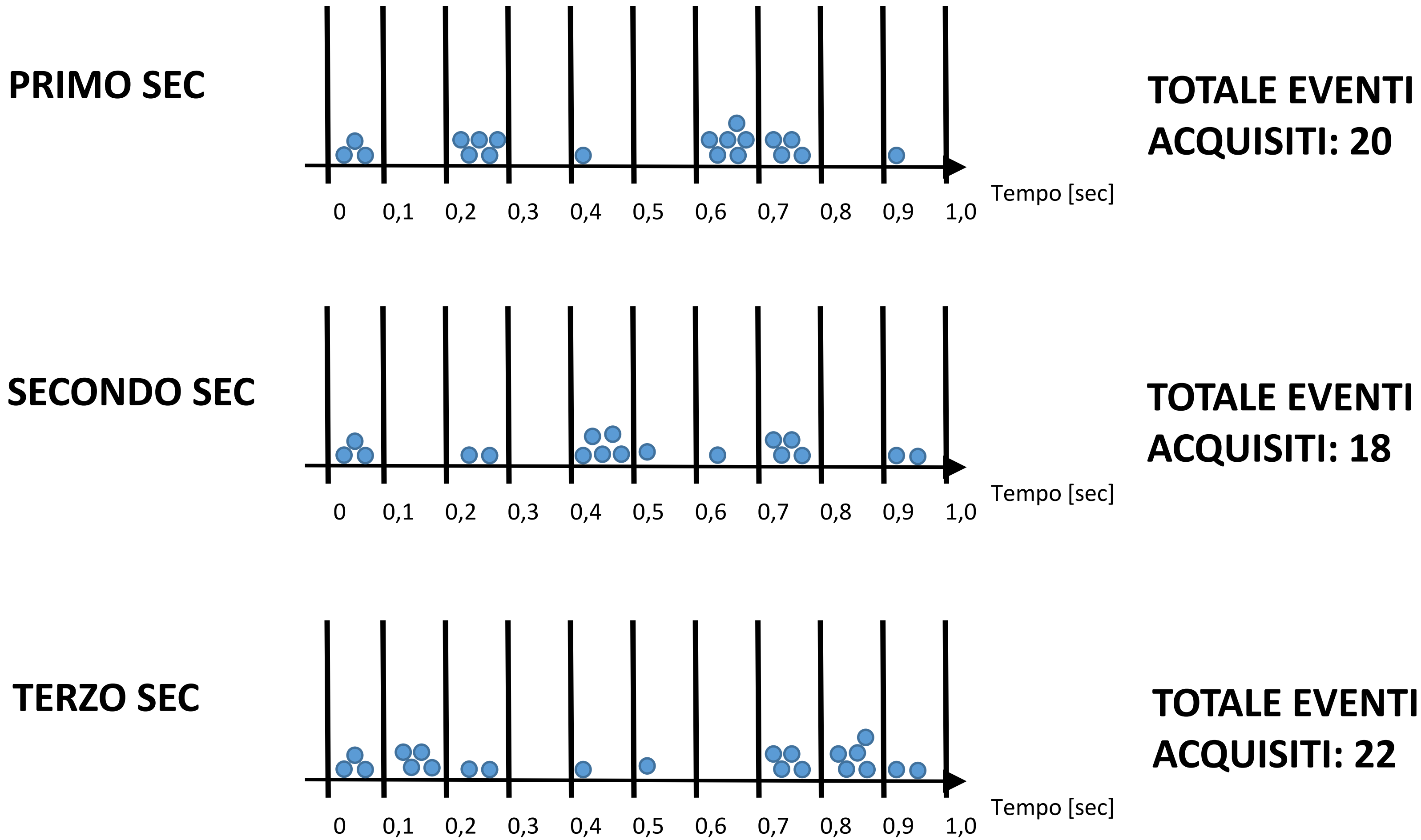
Osservare eventi così rari richiede delle strategie di analisi raffinate, che oggi andremo a scoprire!

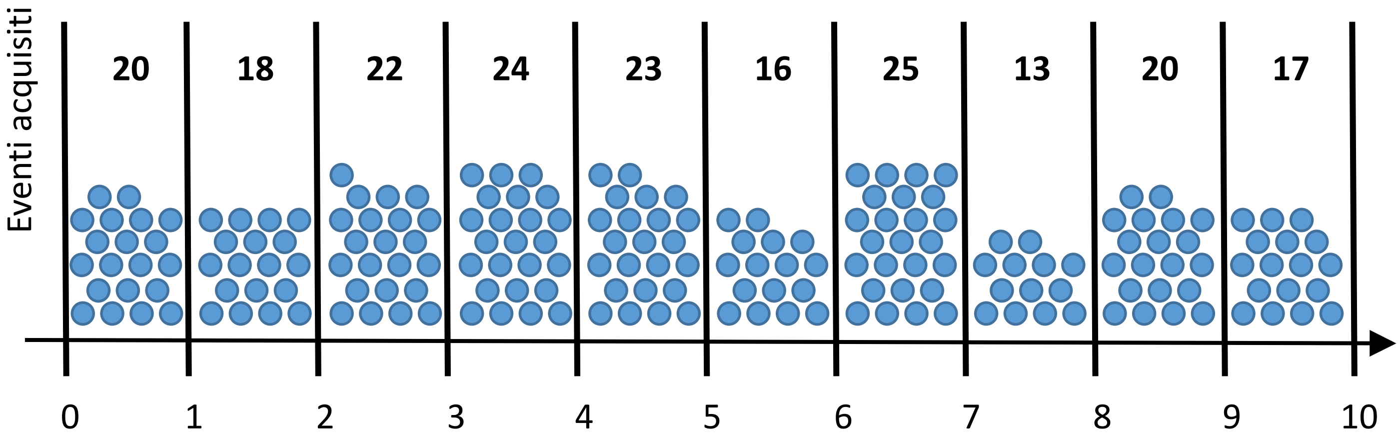
Fluttuazioni statistiche

- Un tipico telescopio di EEE acquisisce eventi con una frequenza media di circa 20 Hz
- Tali eventi sono per la gran parte dovuti a sciami prodotti da raggi cosmici di bassa energia che possono al massimo interessare un singolo telescopio
- Supponendo costante la frequenza di acquisizione di 20 Hz ed ogni evento non correlato al successivo, in un secondo verranno raccolti mediamente 20 eventi distribuiti temporalmente con ugual probabilità lungo l'intervallo di tempo considerato

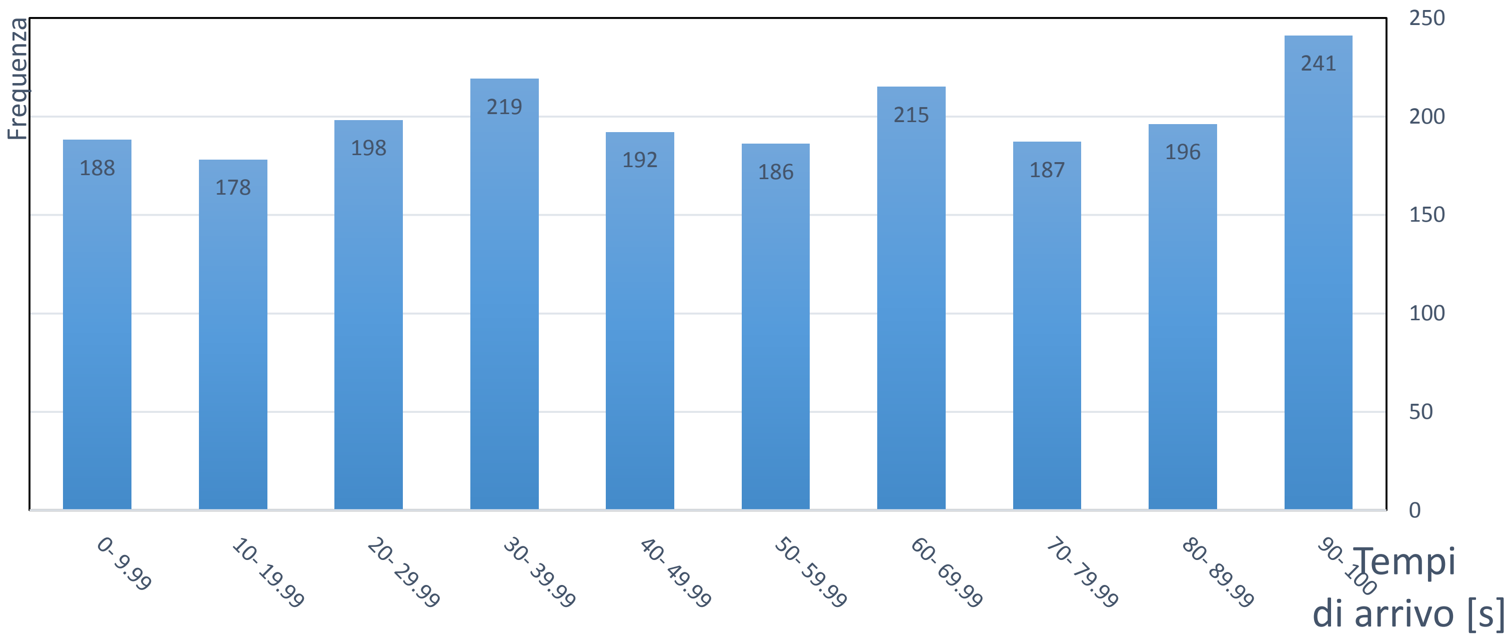


Acquisiamo dati per 3 secondi …





TOTALE EVENTI RACCOLTI: 198
MEDIA AL SECONDO: 19,8



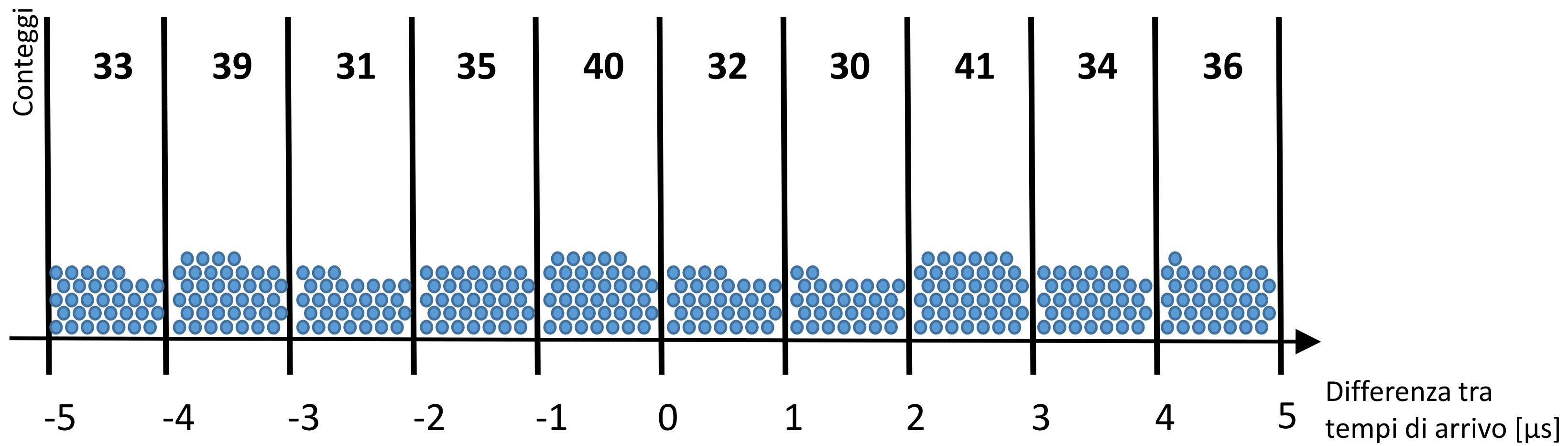
Sincronizzazione tra telescopi

Oltre a registrare i dati provenienti dalle camere il sistema di acquisizione dati registra anche le informazioni generate dal sistema GPS.

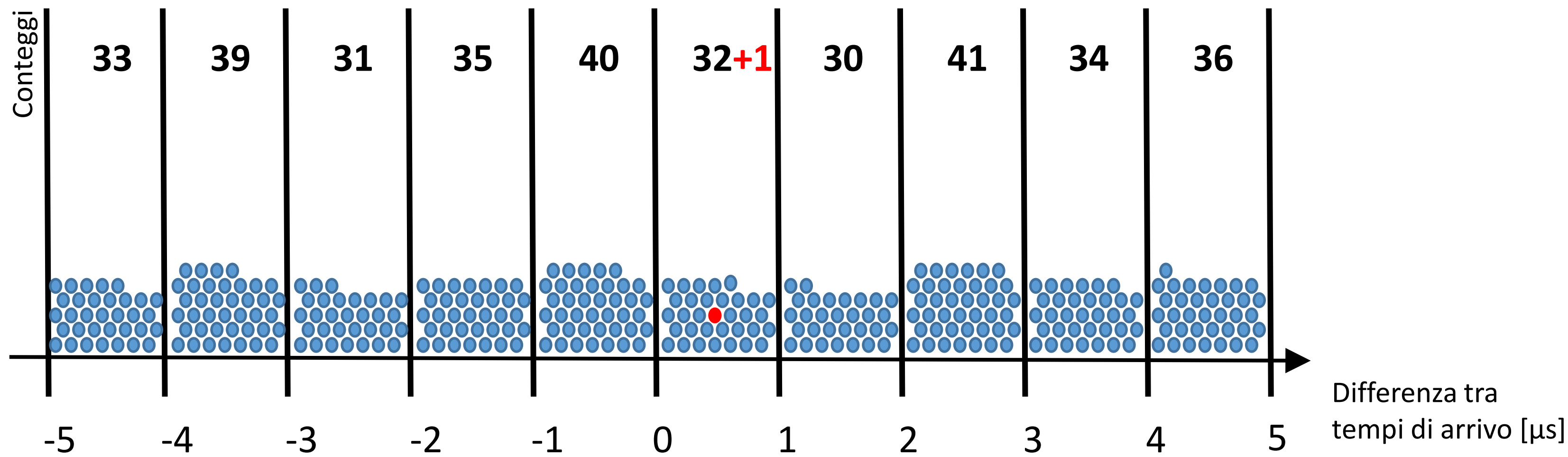
L'utilizzo di un GPS permette la sincronizzazione di più telescopi.

Eventi con lo stesso “etichetta temporale” possono essere analizzati per individuare sciami estesi, ovvero sciami generati da particelle ad altissima energia, oppure sciami correlati (LDC).

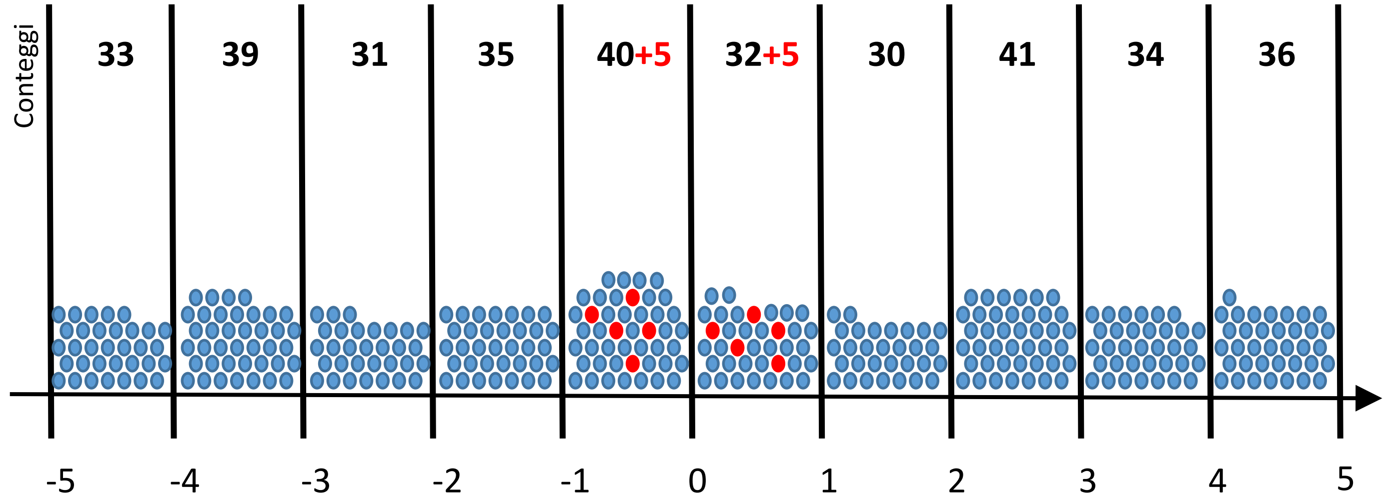
Osservare le coincidenze



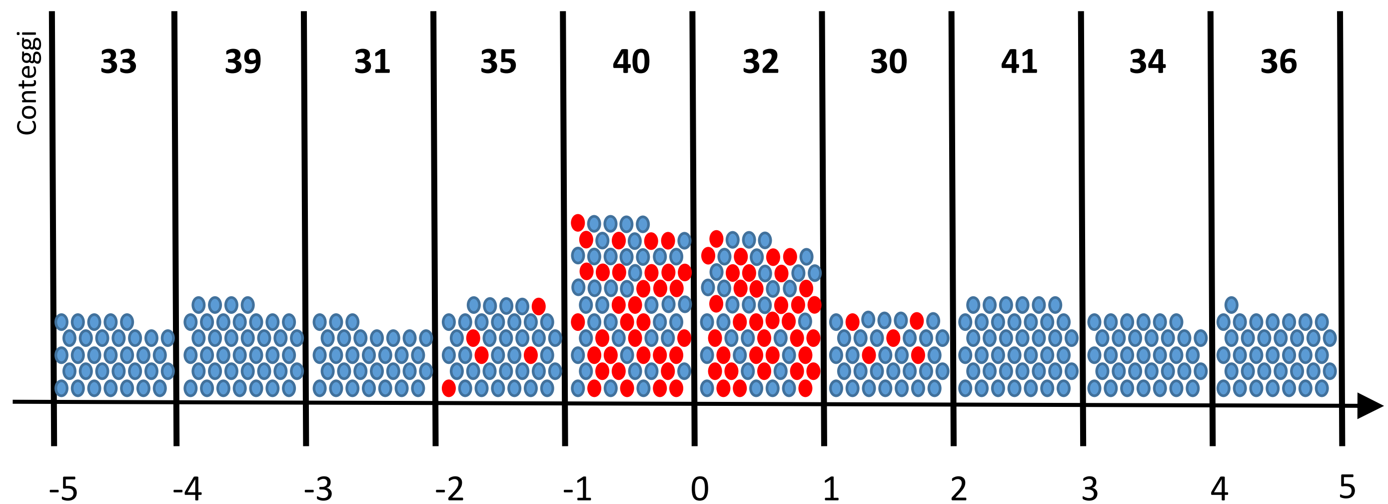
Aggiungiamo 1 conteggio di segnale … indistinguibile!



10 conteggi di segnale sono molto difficilmente rivelabili

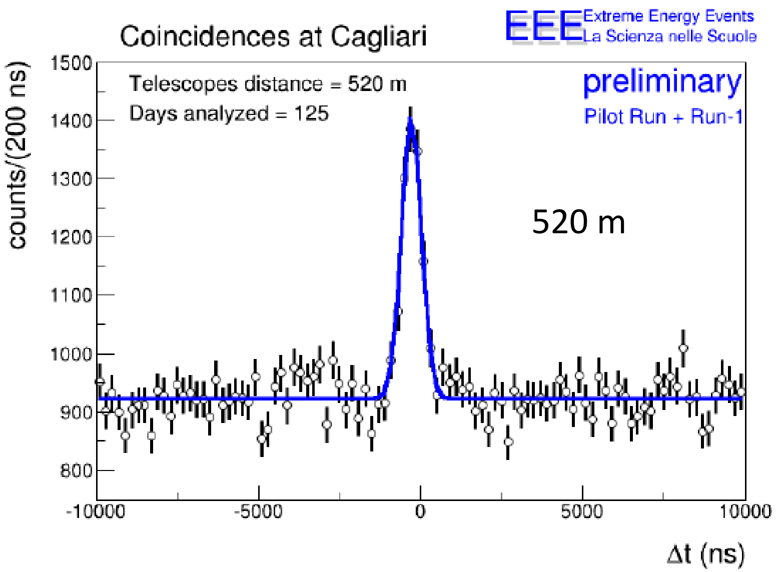


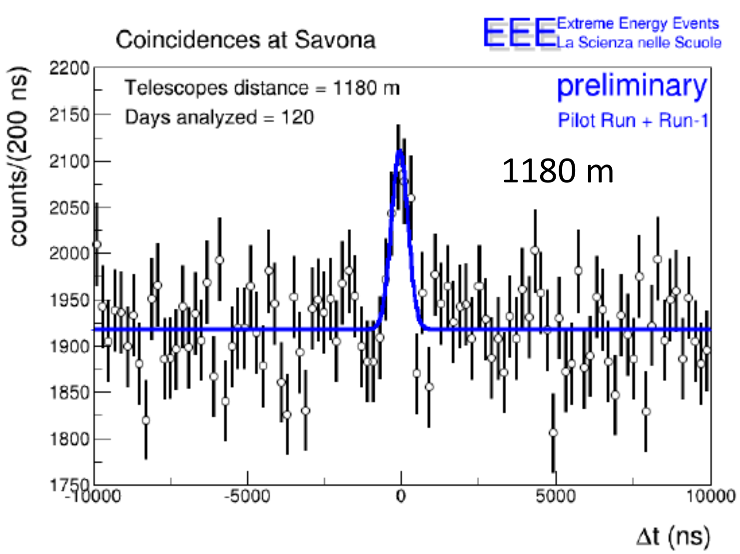
80 conteggi di segnale creano un accumulo ben evidente

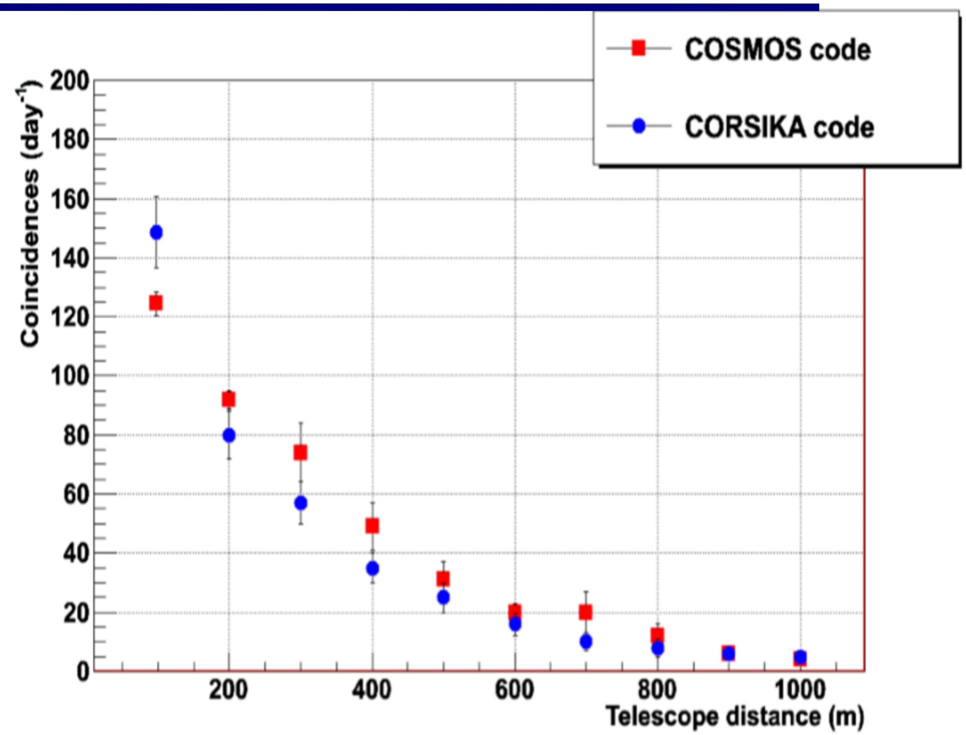


Dobbiamo capire la nostra situazione

Esempio: osservazione di sciami estesi

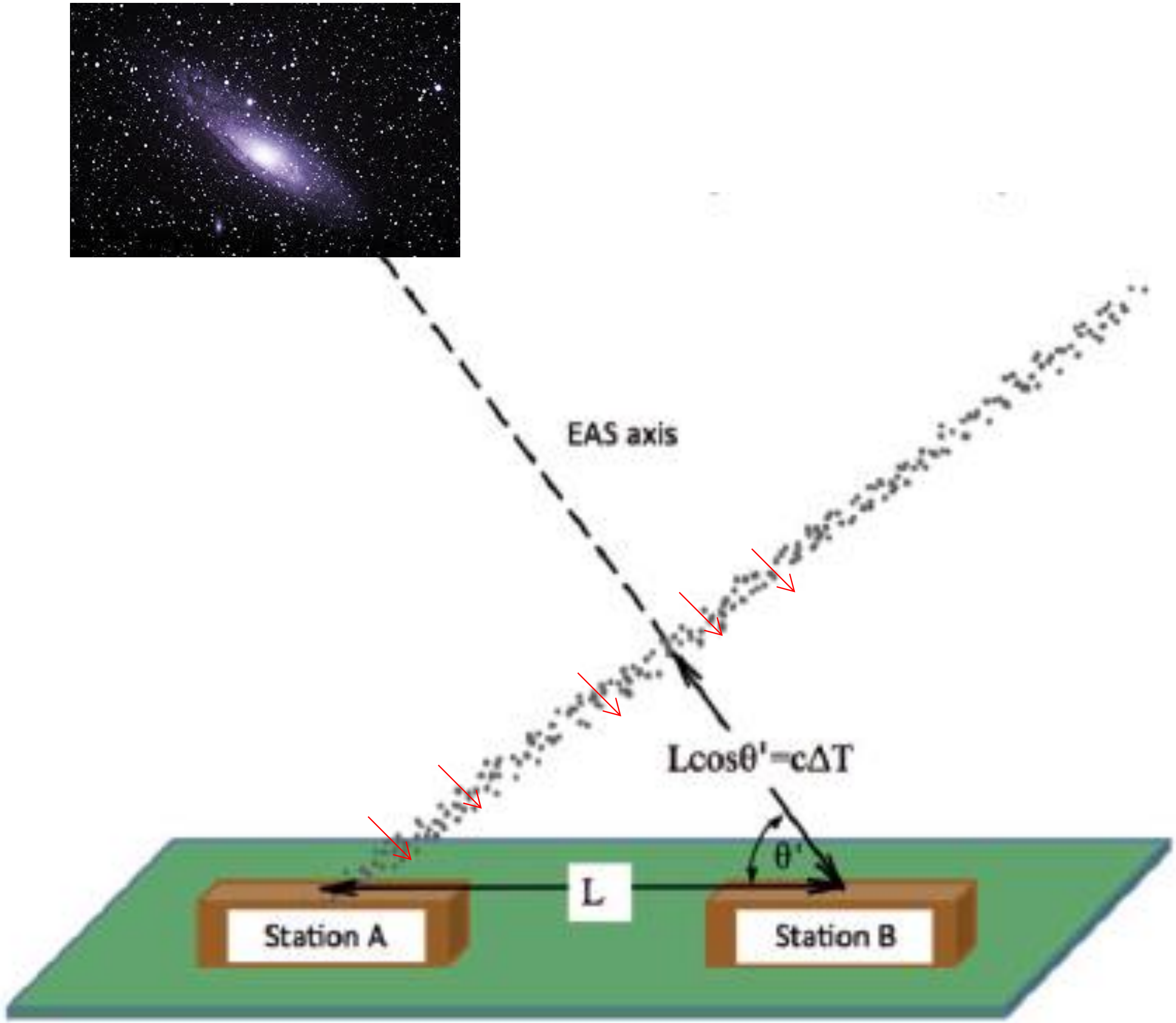






All’aumentare della distanza la probabilità di avere sciami abbastanza estesi da investire entrambe le stazione cala repentinamente, mentre la probabilità di avere coincidenze casuali rimane costante

Larghezza del picco



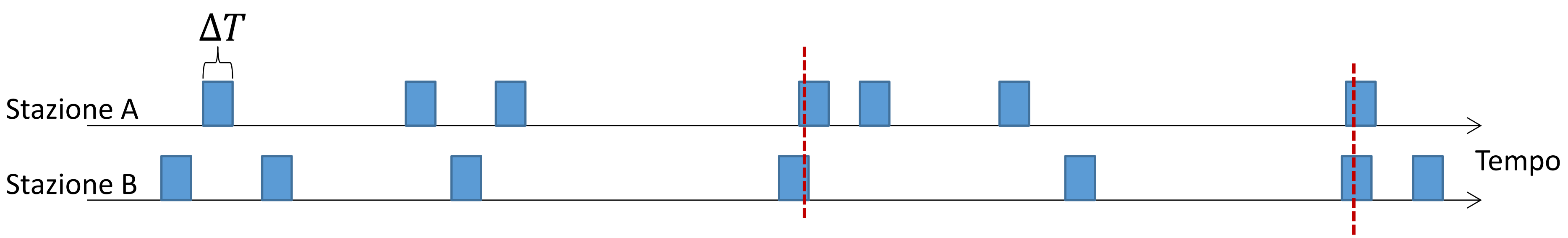
Stimare il fondo

Possiamo stimare qual è la probabilità di avere simultaneamente degli eventi <u>non</u> correlati (non legati a LDC o allo stesso sciame) in due telescopi?

$$f_{acc} \approx 2 * \Delta T f_1 \, f_2 = W f^2$$

f_{acc}= frequenza coincidenze accidentali
ΔT = ampiezza della finestra di osservazione
$f_{1,2}$ = frequenza di acquisizione

Formula generale con largo campo di impiego



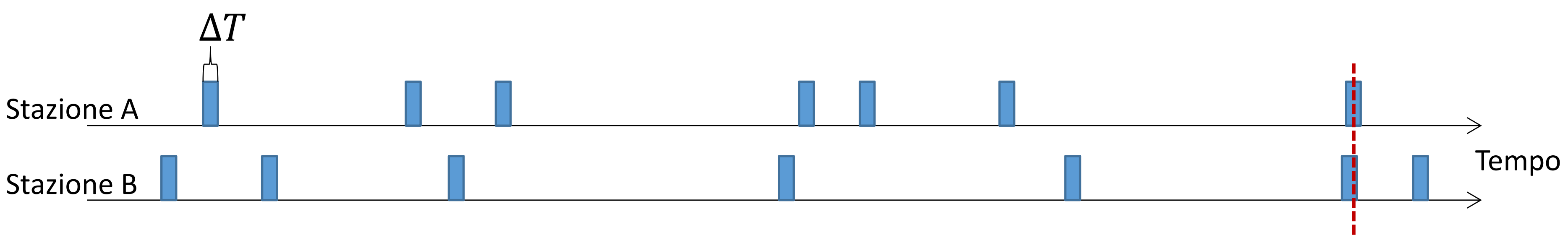
Stimare il fondo

Possiamo stimare qual è la probabilità di avere simultaneamente degli eventi <u>non</u> correlati (non legati a LDC o allo stesso sciame) in due telescopi?

$$f_{acc} \approx 2 * \Delta T f_1\, f_2 = W f^2$$

f_{acc}= frequenza coincidenze accidentali
ΔT = ampiezza della finestra di osservazione
$f_{1,2}$ = frequenza di acquisizione

Formula generale con largo campo di impiego



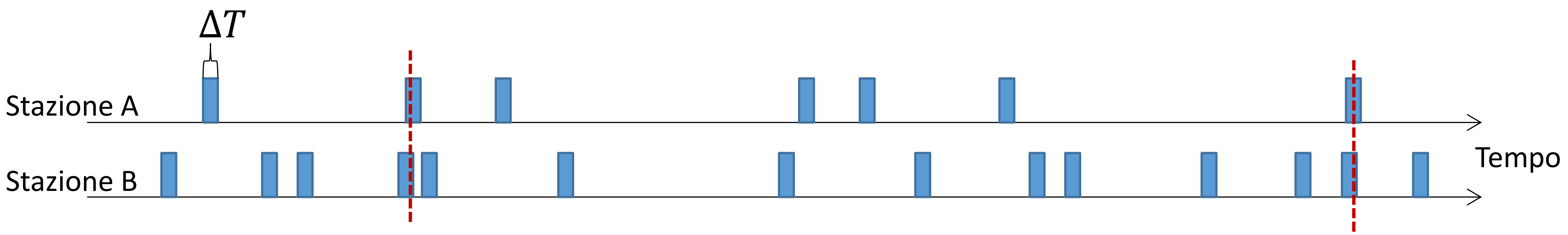
Stimare il fondo

Possiamo stimare qual è la probabilità di avere simultaneamente degli eventi <u>non</u> correlati (non legati a LDC o allo stesso sciame) in due telescopi?

$$f_{acc} \approx 2 * \Delta T f_1 \, f_2 = W f^2$$

f_{acc}= frequenza coincidenze accidentali
ΔT = ampiezza della finestra di osservazione
$f_{1,2}$ = frequenza di acquisizione

Formula generale con largo campo di impiego



Stimare il fondo

Possiamo stimare qual è la probabilità di avere simultaneamente degli eventi non correlati (non legati a LDC o allo stesso sciame) in due telescopi?

$$f_{acc} \approx 2 * \Delta T f_1 \, f_2 = W f^2$$

f_{acc}= frequenza coincidenze accidentali
ΔT = ampiezza della finestra di osservazione
$f_{1,2}$ = frequenza di acquisizione

Formula generale con largo campo di impiego

Risultato: con una finestra di 1 us e una frequenza di acquisizione di ~30 Hz, mi aspetto di avere $f_{acc} \sim 1\ mHz$ → 80 al giorno!

L'osservazione di sciami estesi diventa sempre più complessa al di sopra del km. Ma ancor peggio per le LDC, dove mi aspetto << 1/anno/km^2!!!

Selezionare gli eventi non solo in base al tempo di arrivo

Utilizzare tecniche di analisi più raffinate

E per le LDC?

Anche per le LDC posso richiedere una selezione («taglio») basato sulla direzione delle particelle. Essendo telescopi distanti centinaia di Km devo tener conto anche della curvatura terrestre!

→ Importanza del sistema di riferimento e e della misura di orientazione del telescopio (Ricordate la masterclass di questa mattina !!)

Tuttavia non risulta sufficiente, data la rarità degli eventi che vogliamo misurare:

- Richiesta di più particelle rivelate nello stesso tempo nei telescopi
- Richiesta di coincidenze tra un maggior numero di telescopi → <u>Cluster</u>

Le LDC richiedono finestre ΔT molto superiori da la distanza tra i telescopi e la curvatura terrestre

Station A

$\longleftrightarrow$

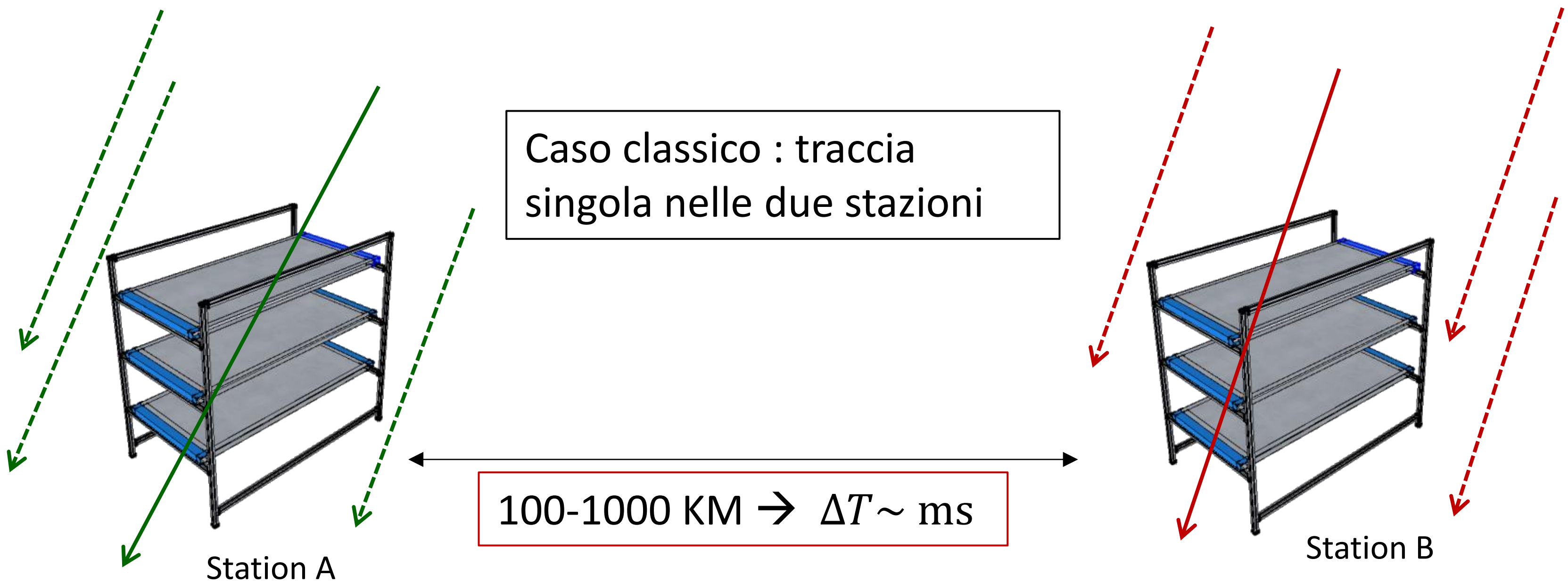
100-1000 KM → $\Delta T \sim \mathrm{ms}$

Station B

E per le LDC?

Tuttavia non risulta sufficiente, data la rarità degli eventi che vogliamo misurare:

- Richiesta di più particelle rivelate nello stesso tempo nei telescopi
- Richiesta di coincidenze tra un maggior numero di telescopi → <u>Cluster</u>



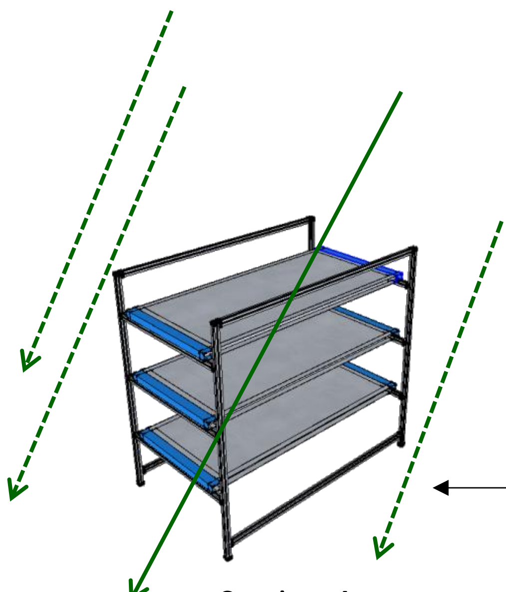
Frequenza attesa LDC: $10^{-3} - 10^{-6}$ / giorno

Frequenza singola stazione: 10-40 Hz → 1HZ → 10^{5} /giorno

E per le LDC?

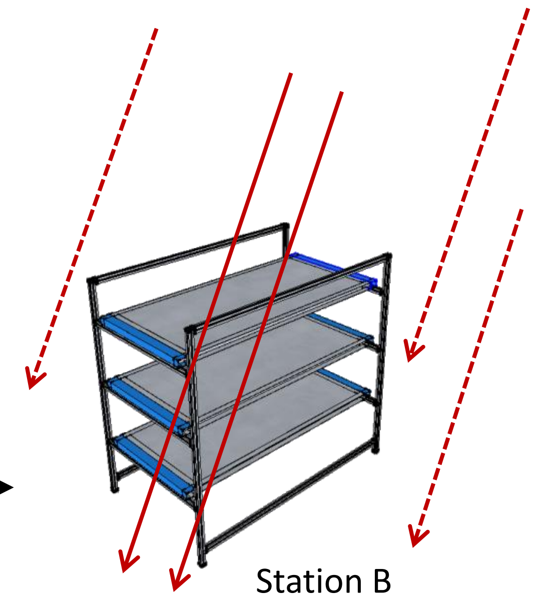
Tuttavia non risulta sufficiente, data la rarità degli eventi che vogliamo misurare:

- Richiesta di più particelle rivelate nello stesso tempo nei telescopi
- Richiesta di coincidenze tra un maggior numero di telescopi → Cluster



Soluzione #1 : traccia singola + doppia traccia

100-1000 KM → $\Delta T \sim \mathrm{ms}$



Frequenza attesa LDC: $10^{-3} - 10^{-6}$ / giorno

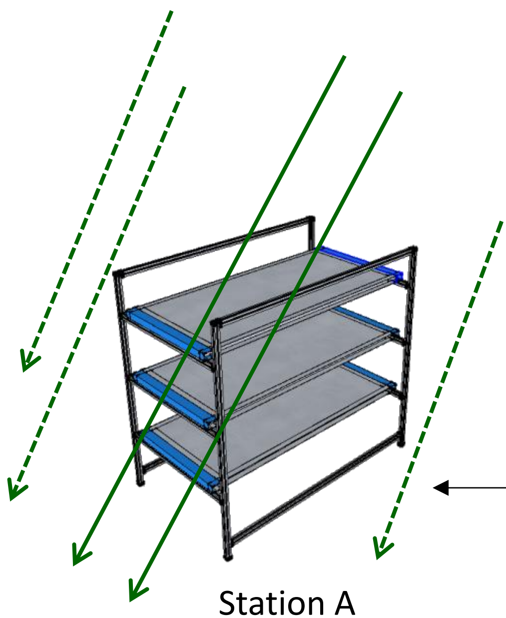
Frequenza singola stazione: 10-40 Hz
Frequenza doppia : 0.01-0.1 Hz

$10^{-4} - 10^{-3}$ HZ →
$100 - 1000$ /giorno

E per le LDC?

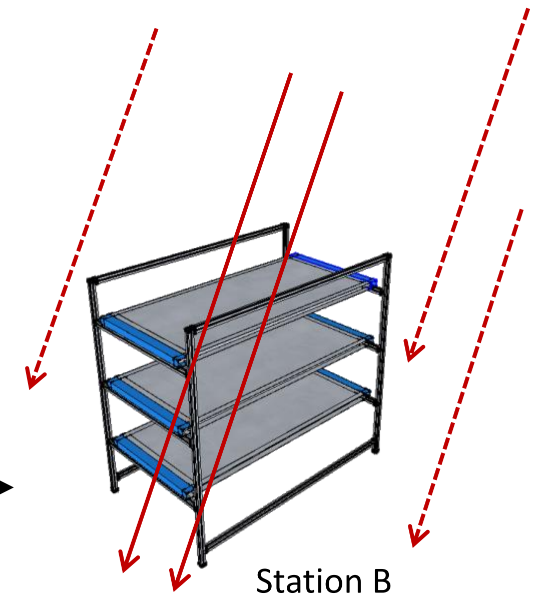
Tuttavia non risulta sufficiente, data la rarità degli eventi che vogliamo misurare:

- Richiesta di più particelle rivelate nello stesso tempo nei telescopi
- Richiesta di coincidenze tra un maggior numero di telescopi → Cluster



Soluzione #2 : doppia traccia in entrambi

100-1000 KM → $\Delta T \sim \mathrm{ms}$



Frequenza attesa LDC: $10^{-3} - 10^{-6}$ / giorno

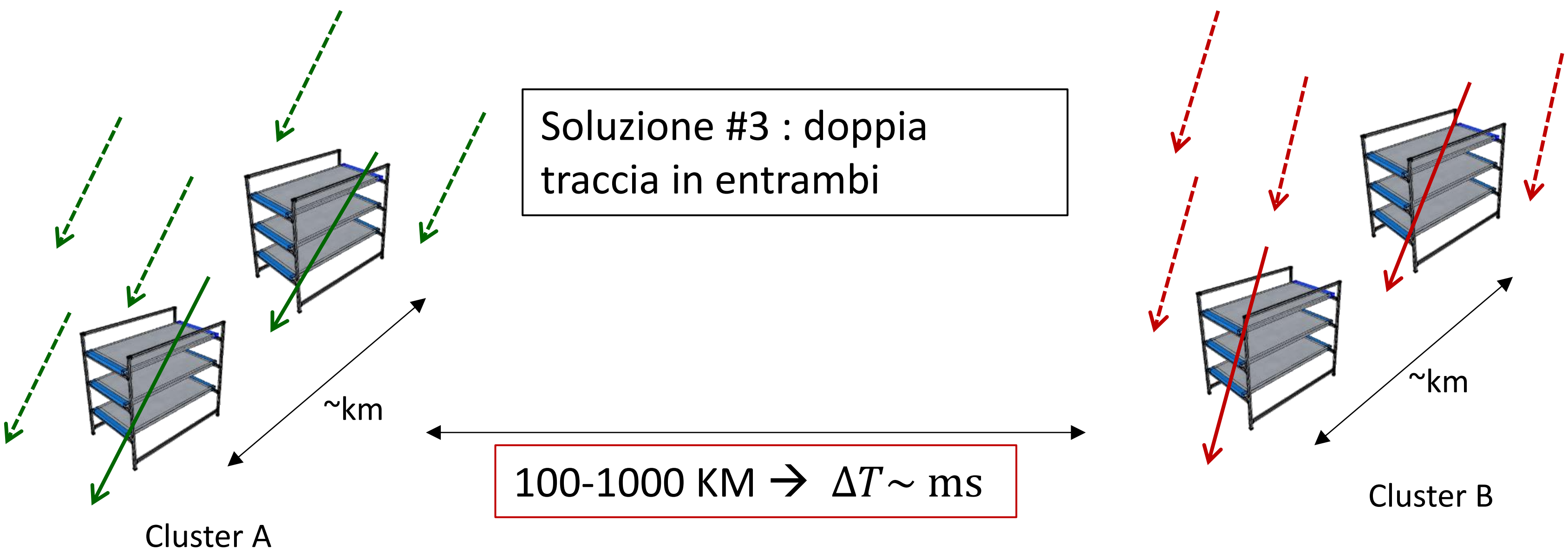
Frequenza singola stazione: 10-40 Hz
Frequenza doppia : 0.01-0.1 Hz

→ $10^{-7} - 10^{-5}$ HZ → $0{,}01 - 1$ /giorno

I cluster

Tuttavia non risulta sufficiente, data la rarità degli eventi che vogliamo misurare:

- Richiesta di più particelle rivelate nello stesso tempo nei telescopi
- Richiesta di coincidenze tra un maggior numero di telescopi → Cluster

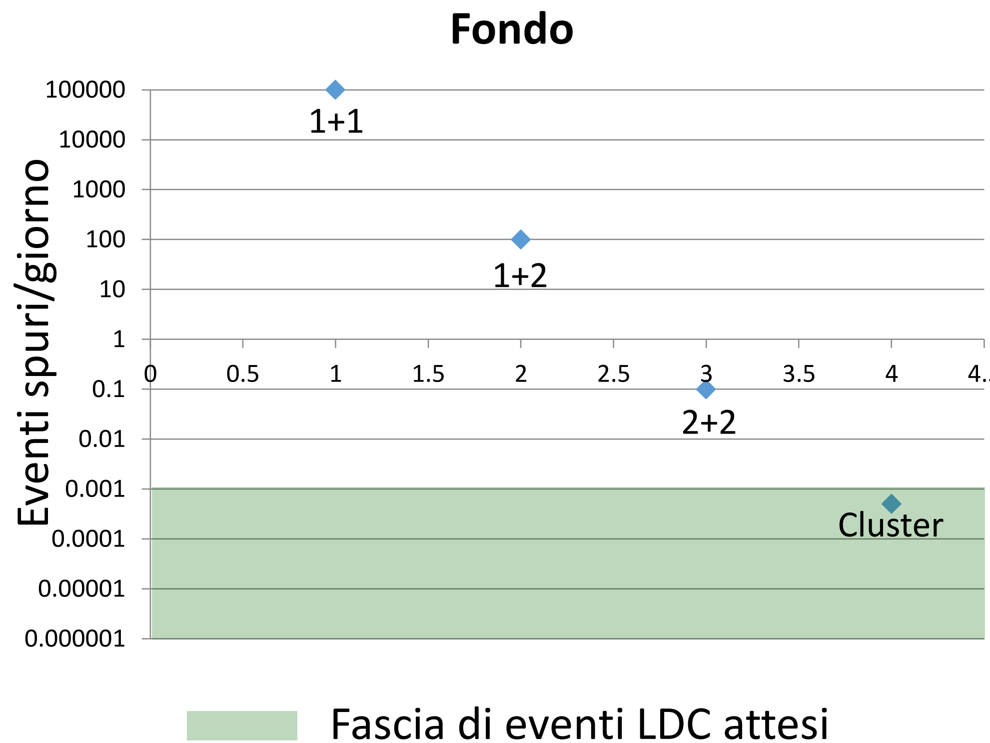


Frequenza attesa LDC: $10^{-3}\ -10^{-6}\ /\ \mathrm{giorno}$

Frequenza Cluster: 0.004-0.04 Hz
Dipende da molteplici fattori

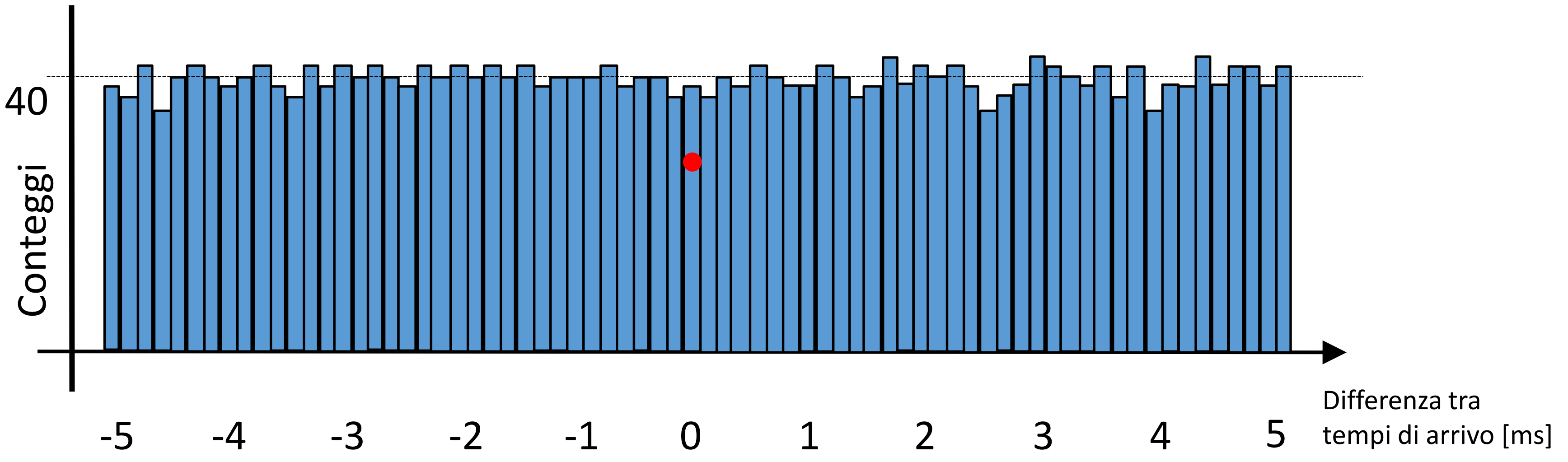
$10^{-8}\ -10^{-7}$ HZ →
$10^{-3} - 10^{-2}/\mathrm{giorno}$

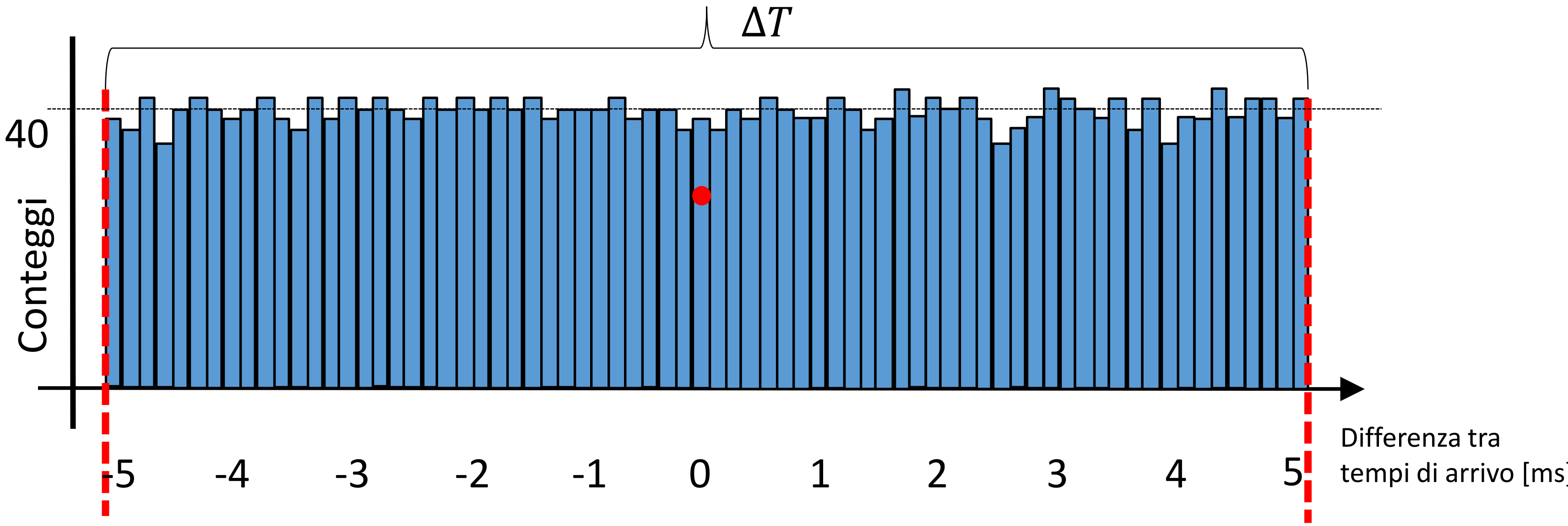
Cluster e LDC



La strategia dei Cluster ci permette di raggiungere le stesse frequenze del segnale da studiare, ma da sola non basta!
Nel caso migliore siamo nella situazione in cui palline blue/rosse sono nello stesso numero
Dobbiamo sviluppare un'opportuna strategia di analisi

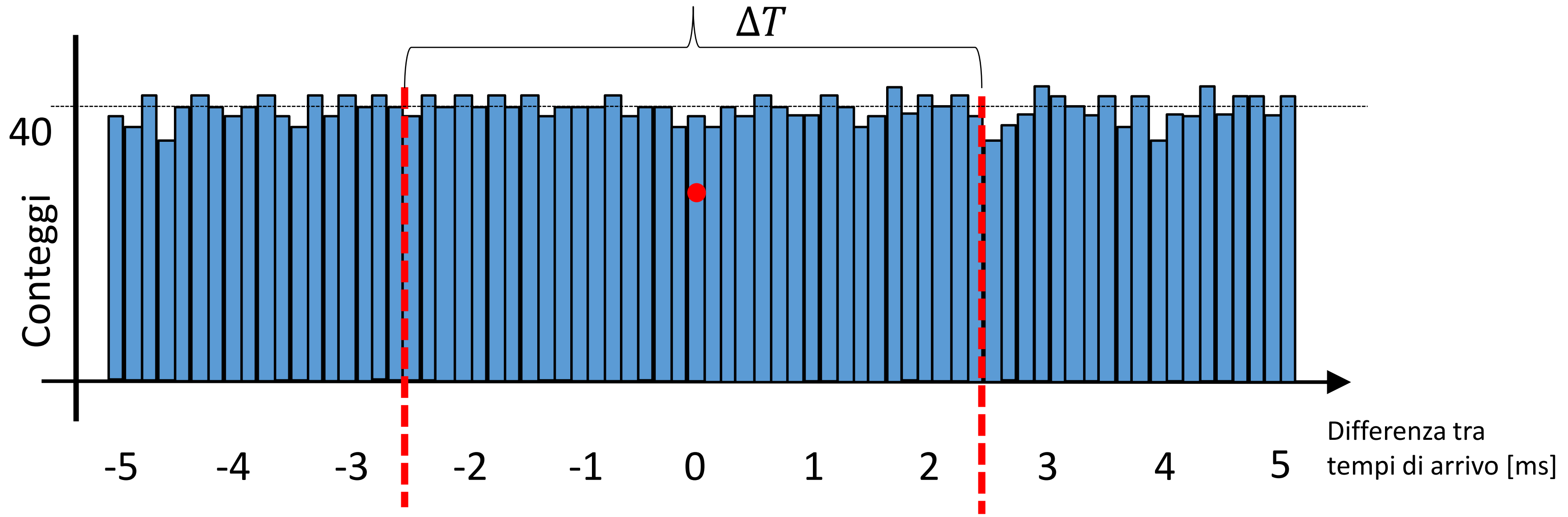
L'analisi

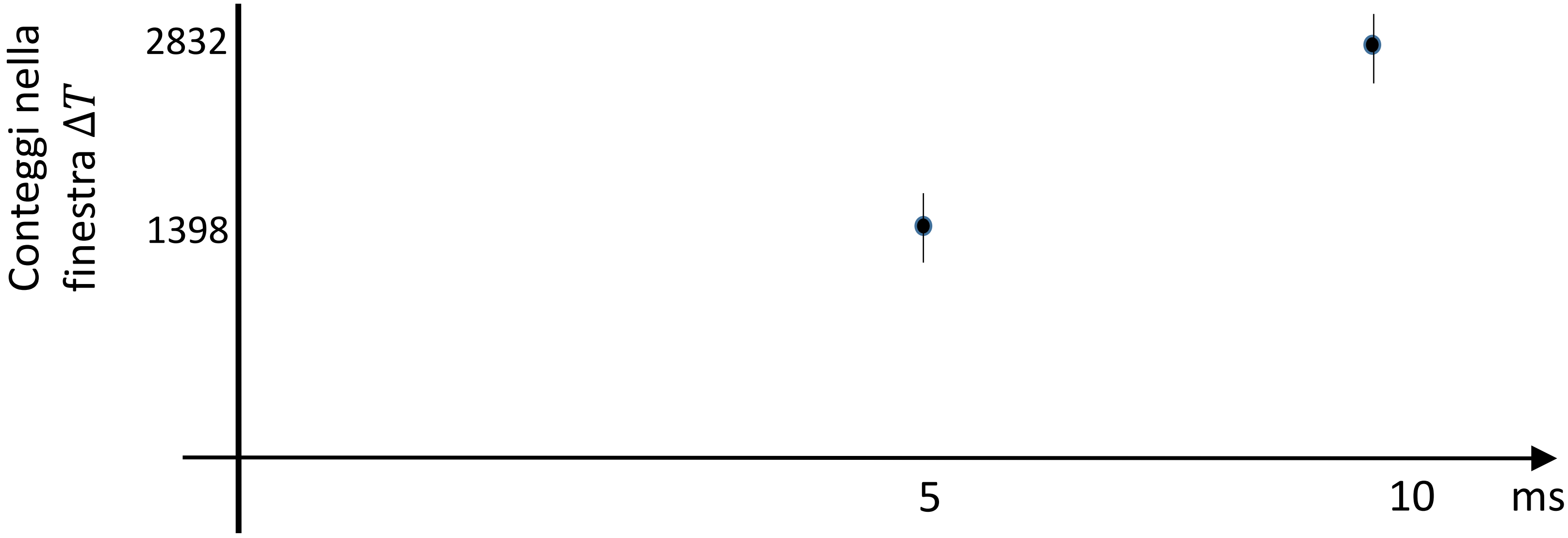


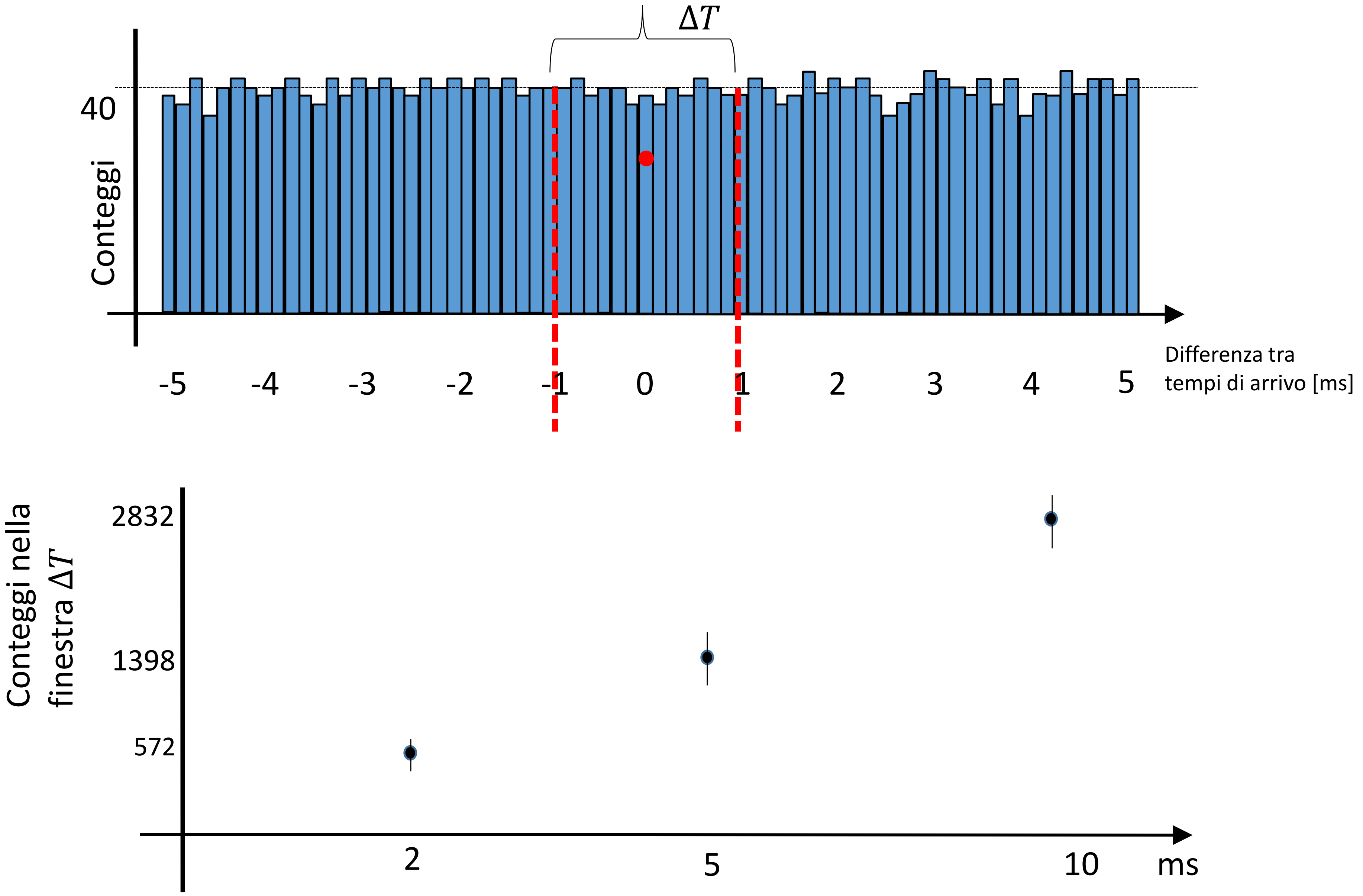


ΔT
40
Conteggi
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differenza tra tempi di arrivo [ms]

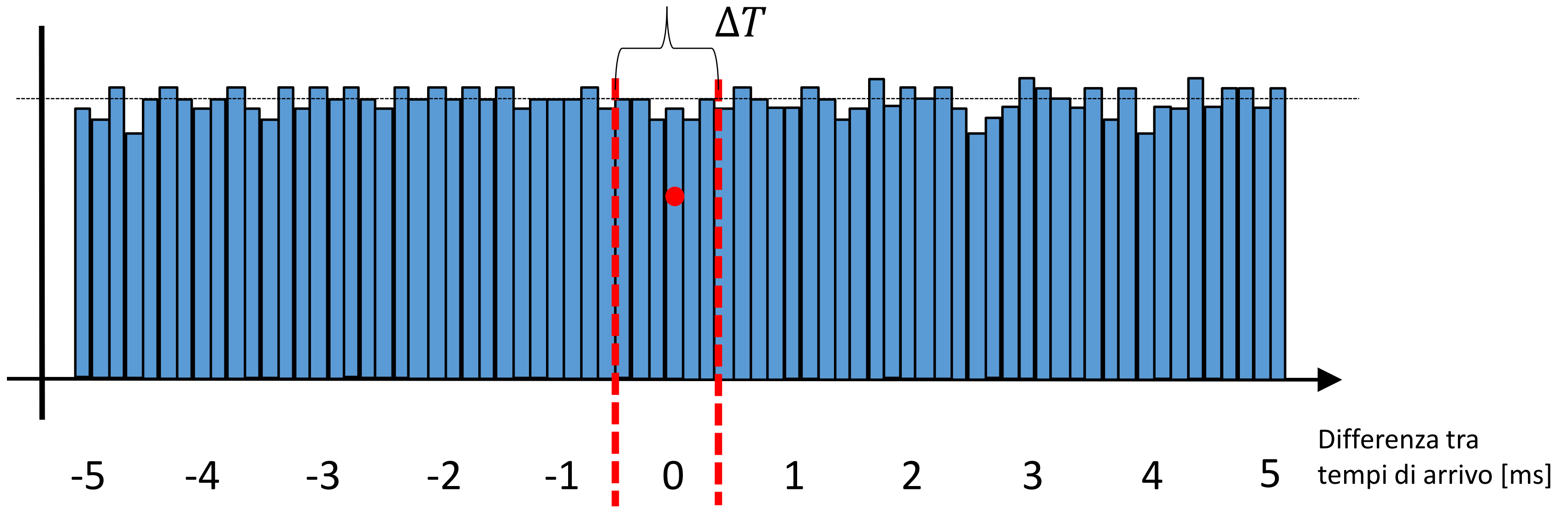
Conteggi nella finestra ΔT
2832
10
ms

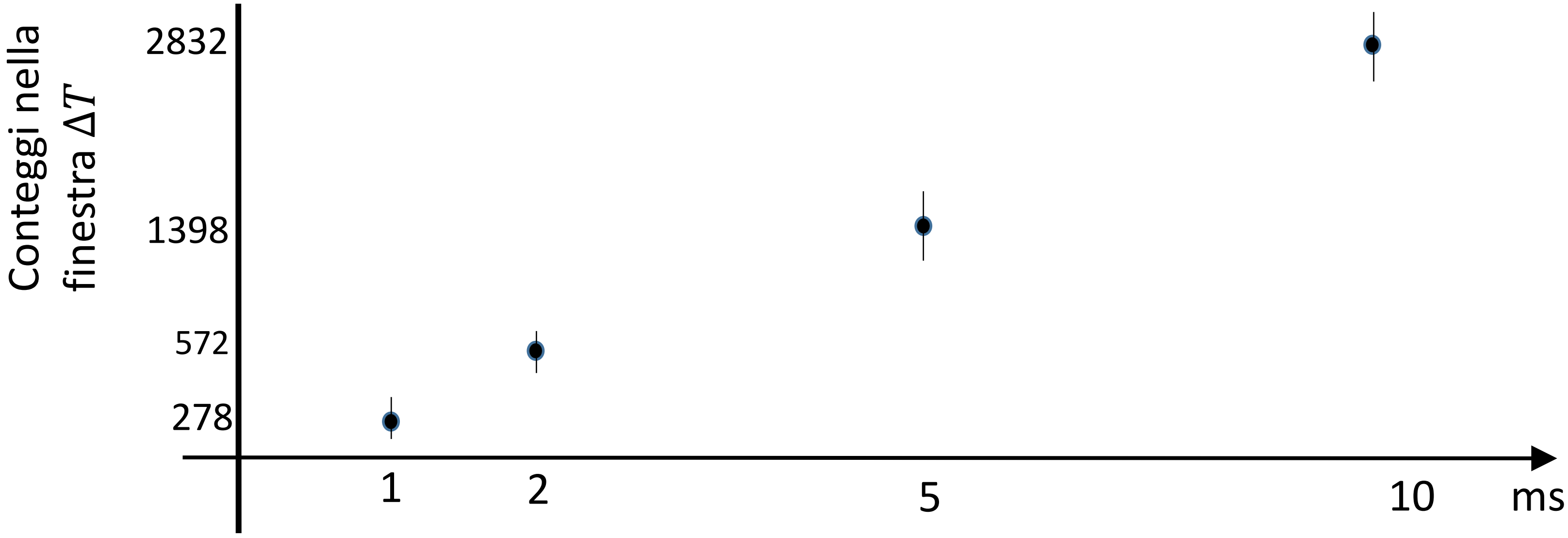

ΔT
40
Conteggi
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differenza tra
tempi di arrivo [ms]


Conteggi nella
finestra ΔT
2832
1398
5
10
ms

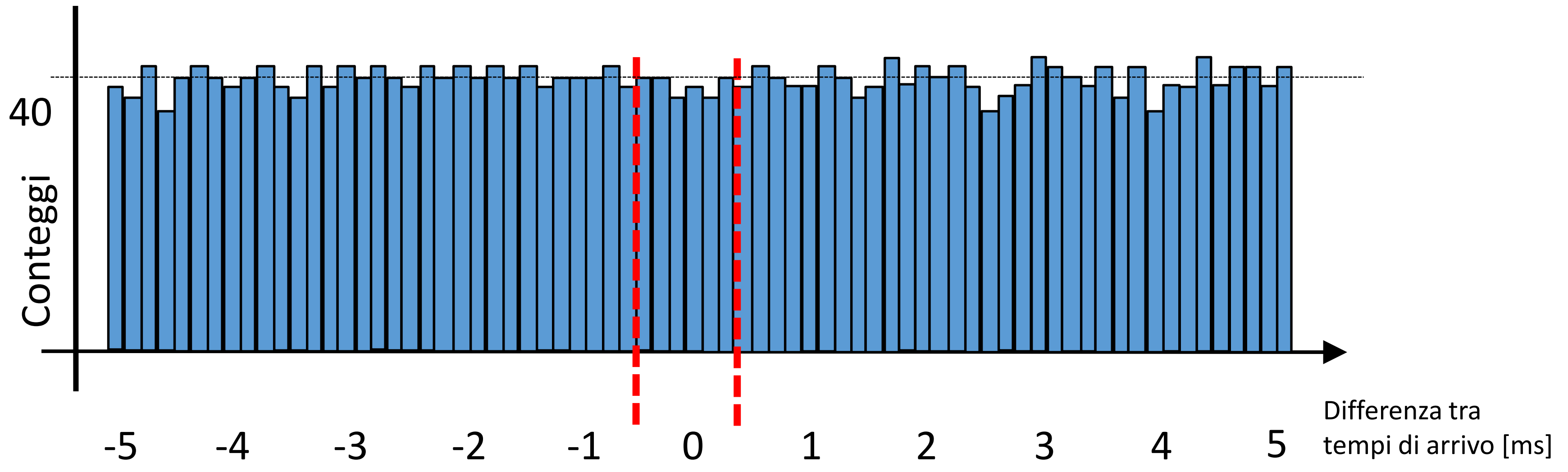


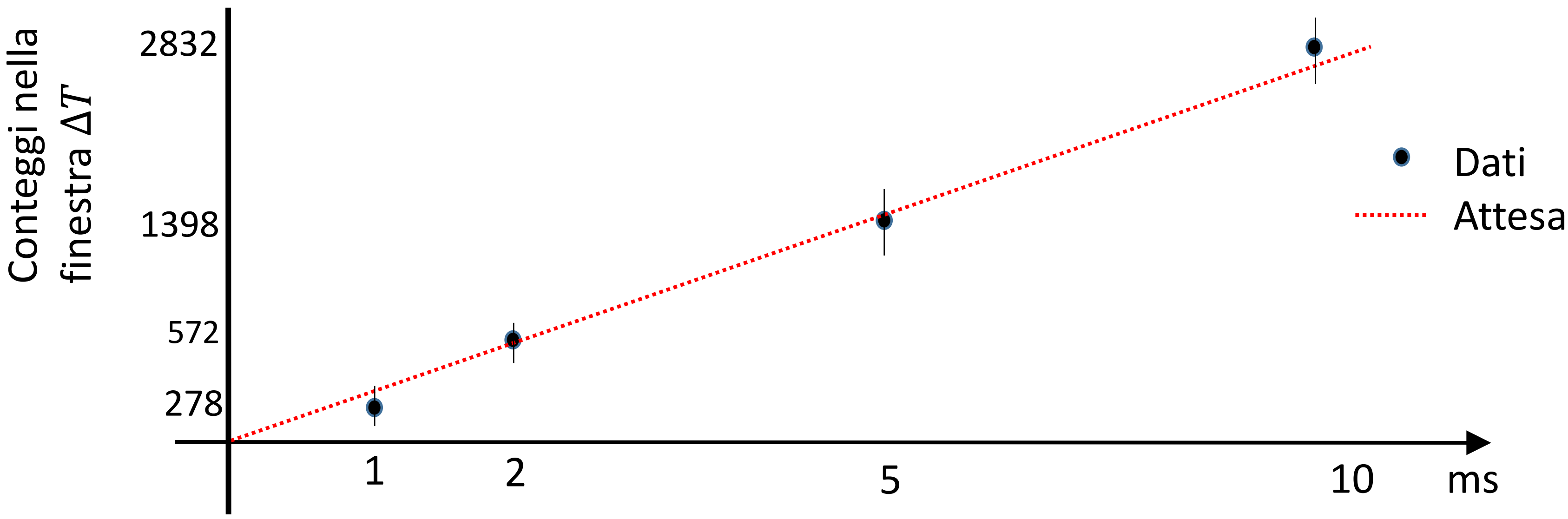
ΔT
40
Conteggi
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differenza tra
tempi di arrivo [ms]
Conteggi nella
finestra ΔT
2832
1398
572
2
5
10
ms

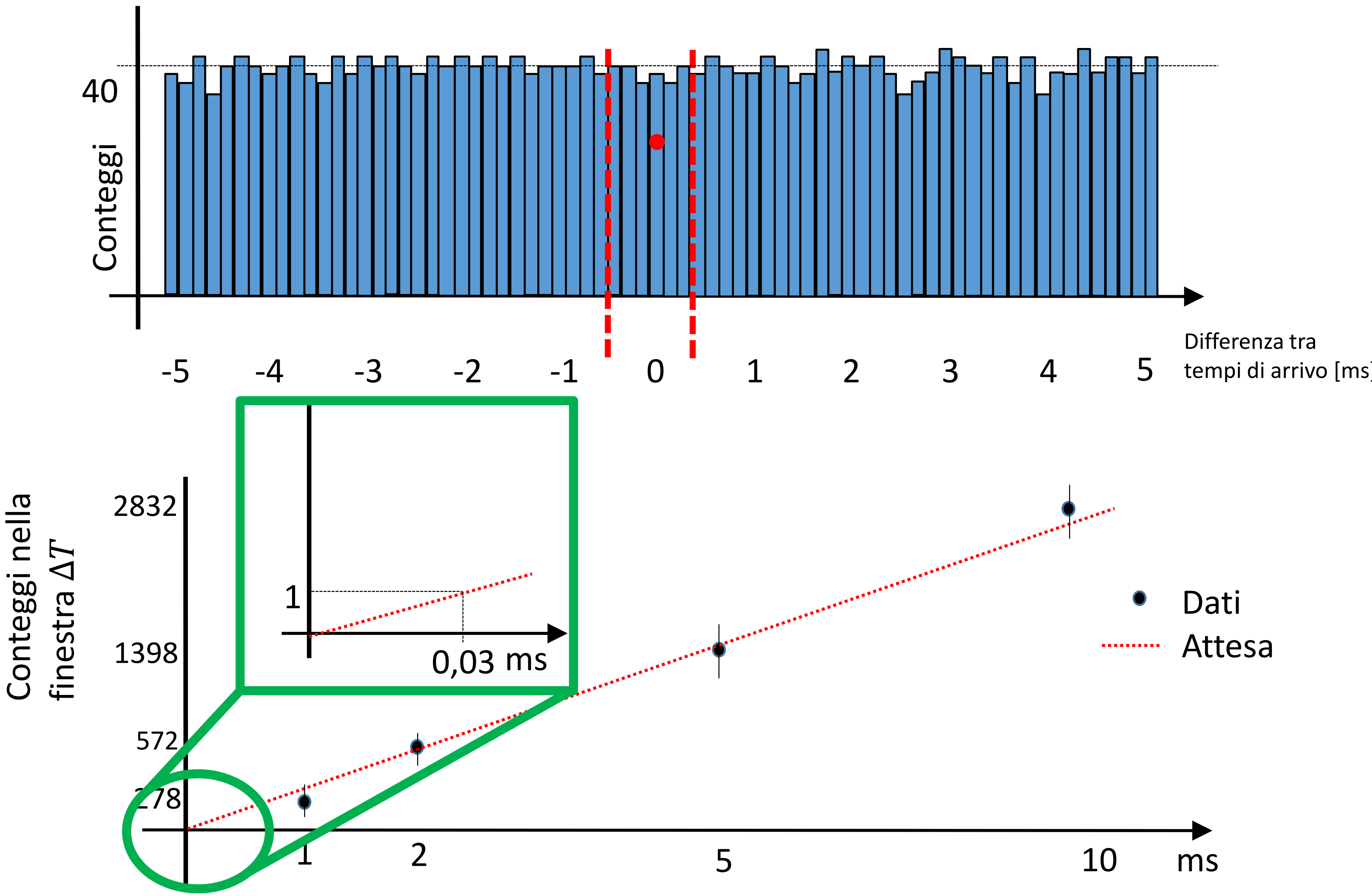

ΔT
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differenza tra
tempi di arrivo [ms]


Conteggi nella
finestra ΔT
2832
1398
572
278
1
2
5
10
ms

L'estrapolazione

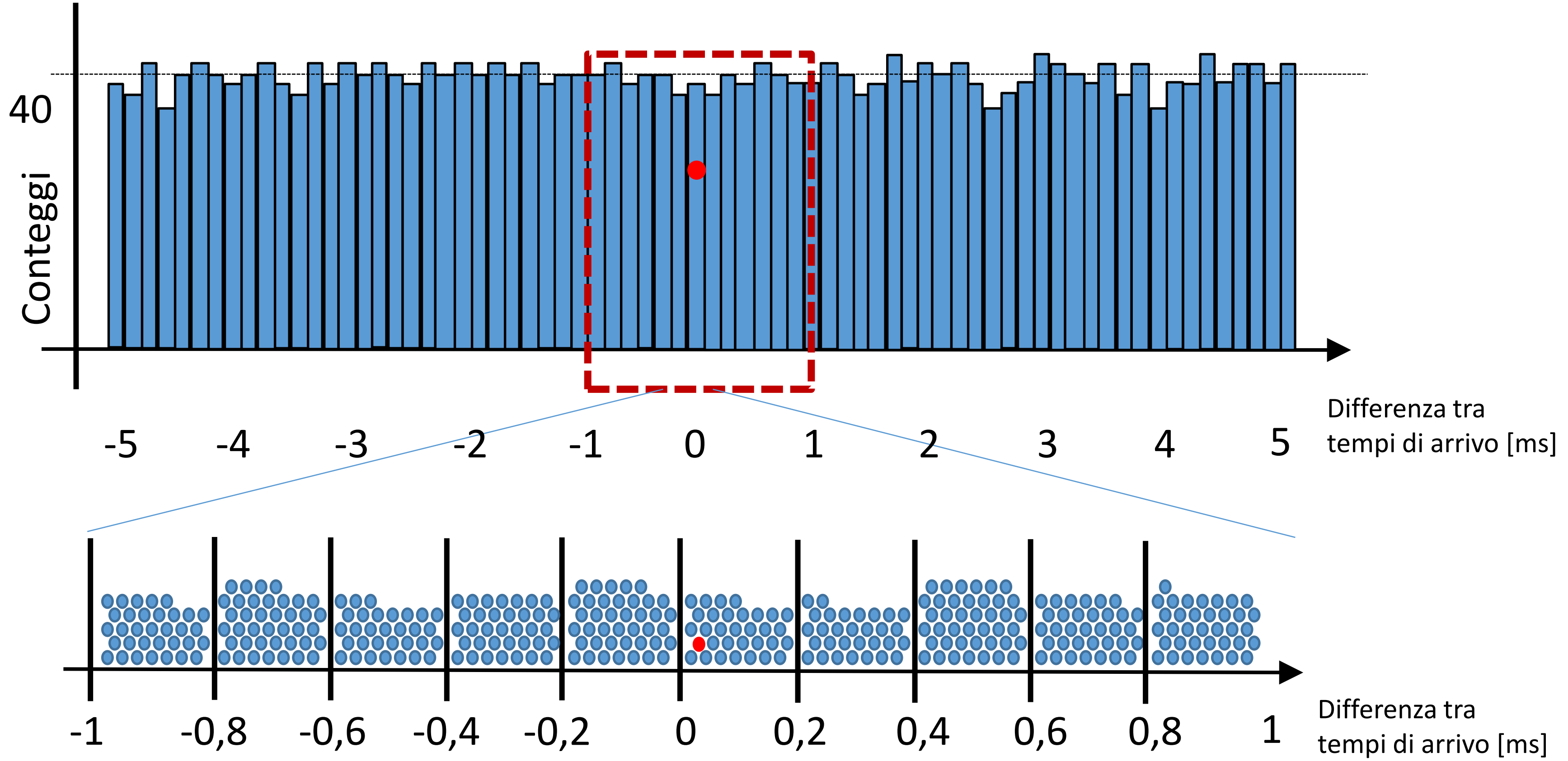


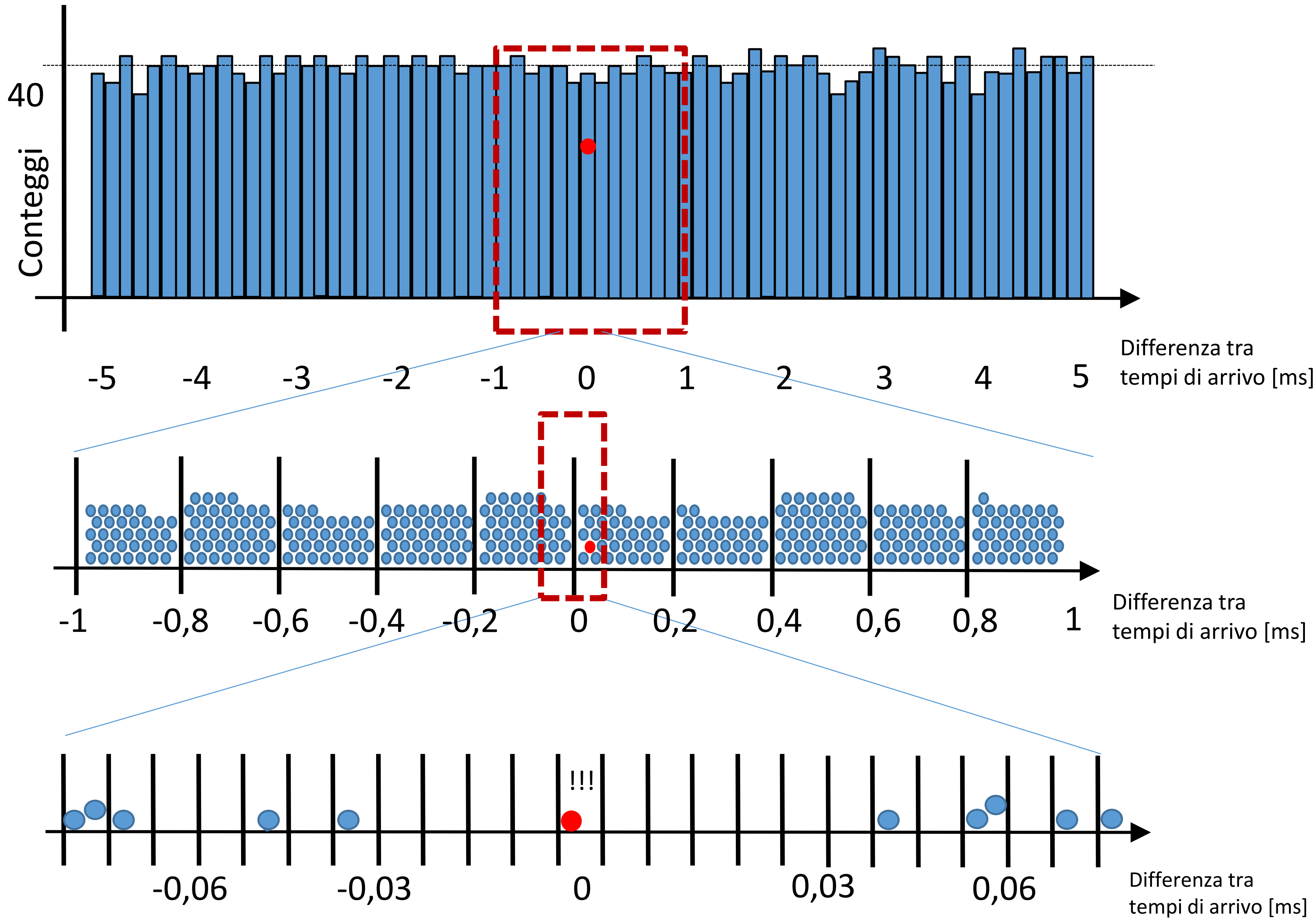




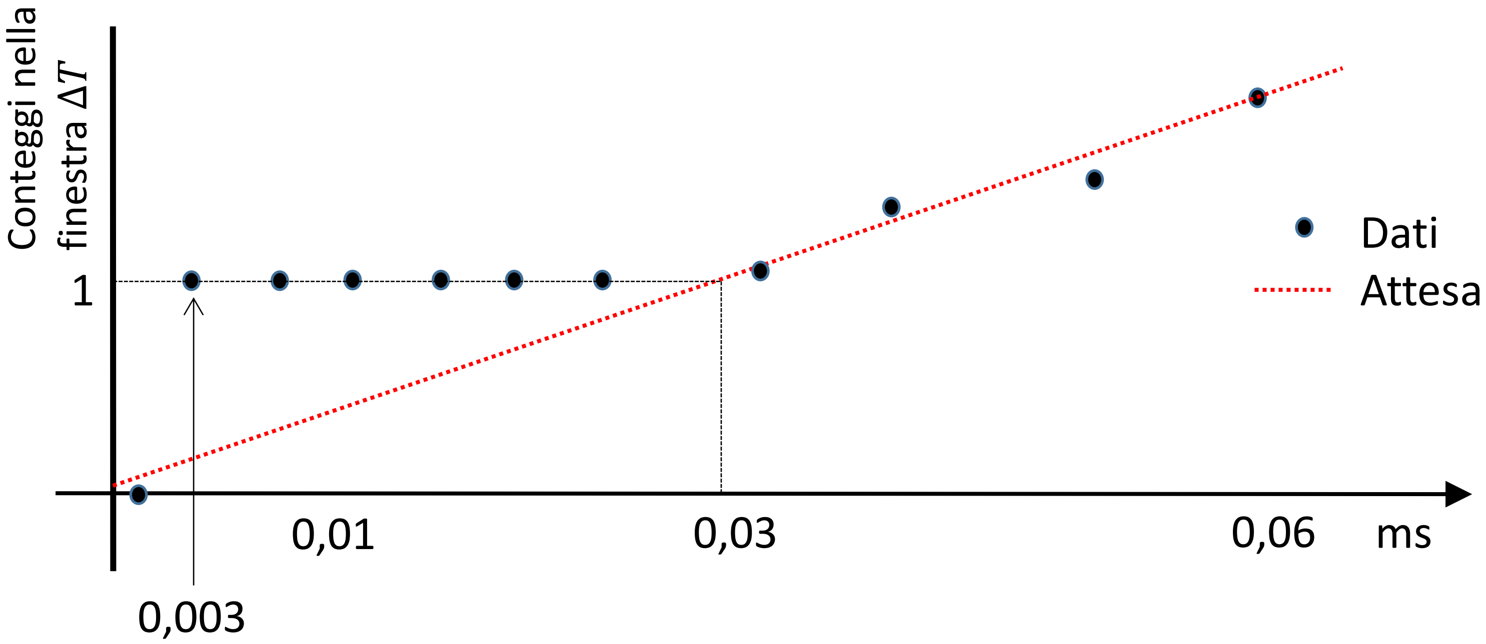
40
Conteggi
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differenza tra
tempi di arrivo [ms]
Conteggi nella
finestra ΔT
2832
1398
572
278
1
0,03 ms
1
2
5
10
ms
Dati
Attesa

Confronto con i dati

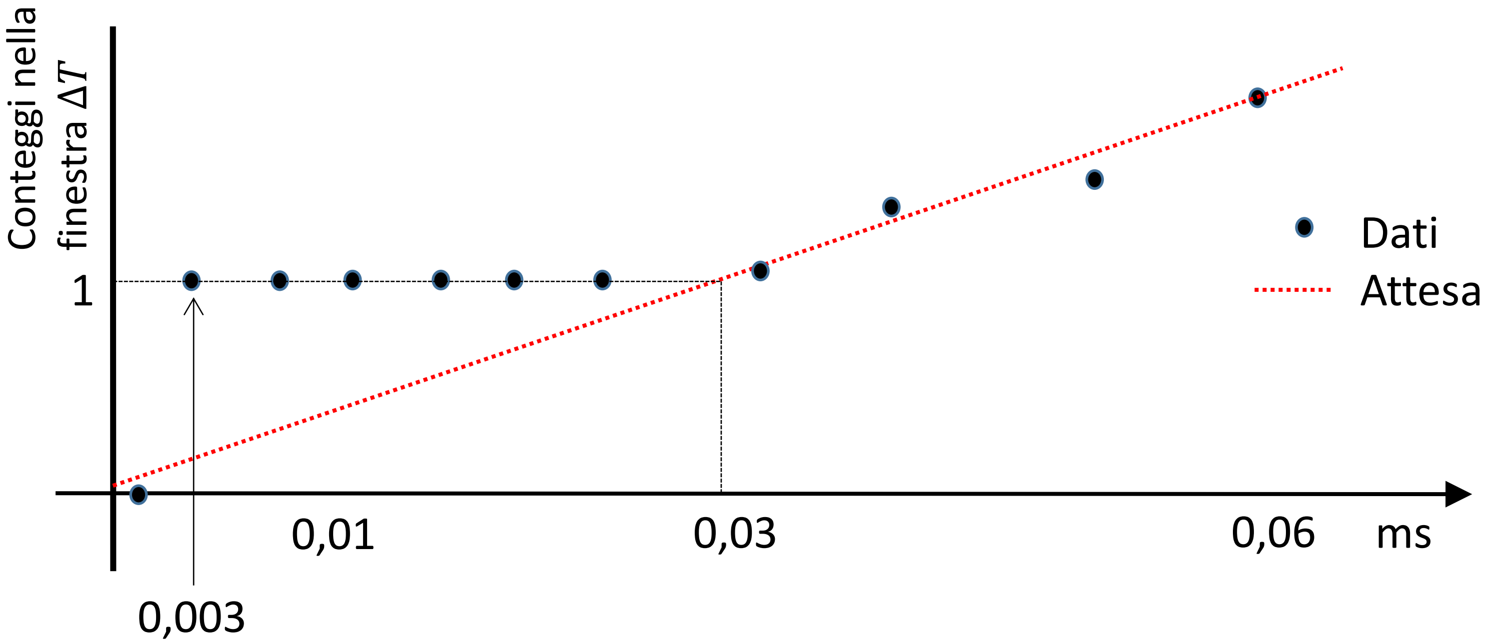



40
Conteggi
-5
-4
-3
-2
-1
0
1
2
3
4
5
Differenza tra tempi di arrivo [ms]
-1
-0,8
-0,6
-0,4
-0,2
0
0,2
0,4
0,6
0,8
1
Differenza tra tempi di arrivo [ms]
!!!
-0,06
-0,03
0
0,03
0,06
Differenza tra tempi di arrivo [ms]

Deviazione dal modello?



Deviazione dal modello?



Un ultimo sforzo! Dobbiamo quantizzare questa deviazione osservata, darle una valenza oggettiva

Il p-value e il test-ipotesi

Il test-ipotesi è una procedura statistica per verificare con quale significatività una misura si discosta dal modello proposto.

1. Si definisce l'ipotesi H_0. Il resto del metodo sarà svolto a dimostrare che H_0 è falsa
2. Si definisce la statistica che descrive il modello
3. Si osserva se e quanto la misura effettuata è improbabile data l'ipotesi H_0, calcolando la probabilità di ottenere un risultato uguale o ancora più lontano dall'ipotesi H_0 (p-value).

Il p-value e il test-ipotesi

Esempio: voglio assicurarmi di non essere imbrogliato sul peso della scatola dei biscotti, venduta per 100±2 gr

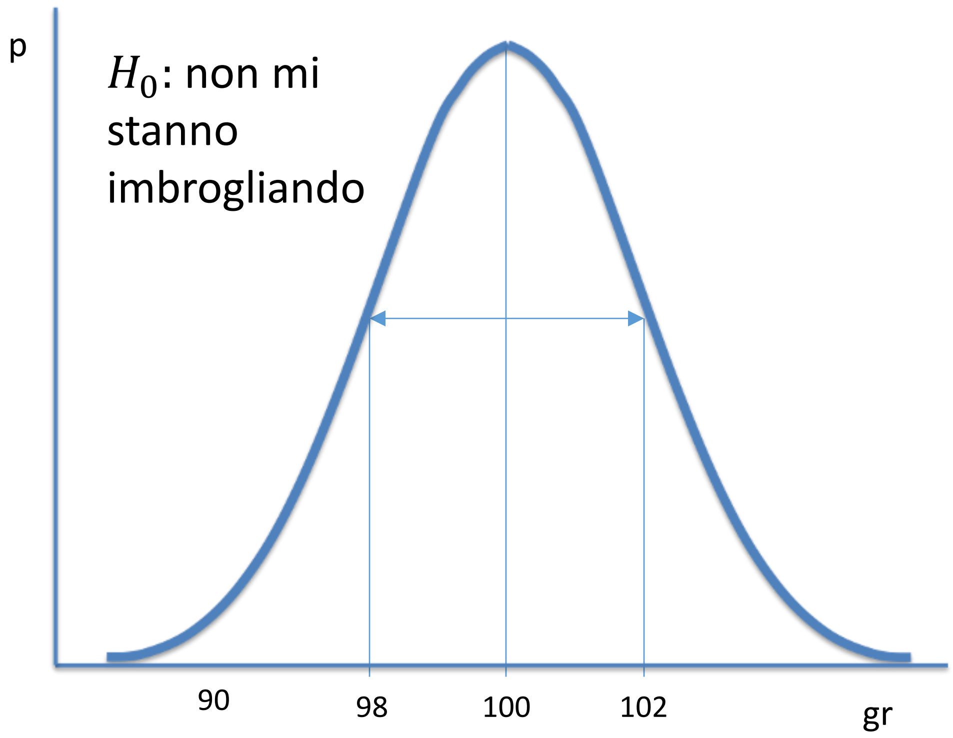
Il test-ipotesi è una procedura statistica per verificare con quale significatività una misura si discosta dal modello proposto.

1. Si definisce l'ipotesi H_0. Il resto del metodo sarà svolto a dimostrare che H_0 è falsa
2. Si definisce la statistica che descrive il modello
3. Si osserva se e quanto la misura effettuata è improbabile data l'ipotesi H_0, calcolando la probabilità di ottenere un risultato uguale o ancora più lontano dall'ipotesi H_0 (p-value).

Il p-value e il test-ipotesi

Esempio: voglio assicurarmi di non essere imbrogliato sul peso della scatola dei biscotti, venduta per 100±2 gr

Il test-ipotesi è una procedura statistica per verificare con quale significatività una misura si discosta dal modello proposto.

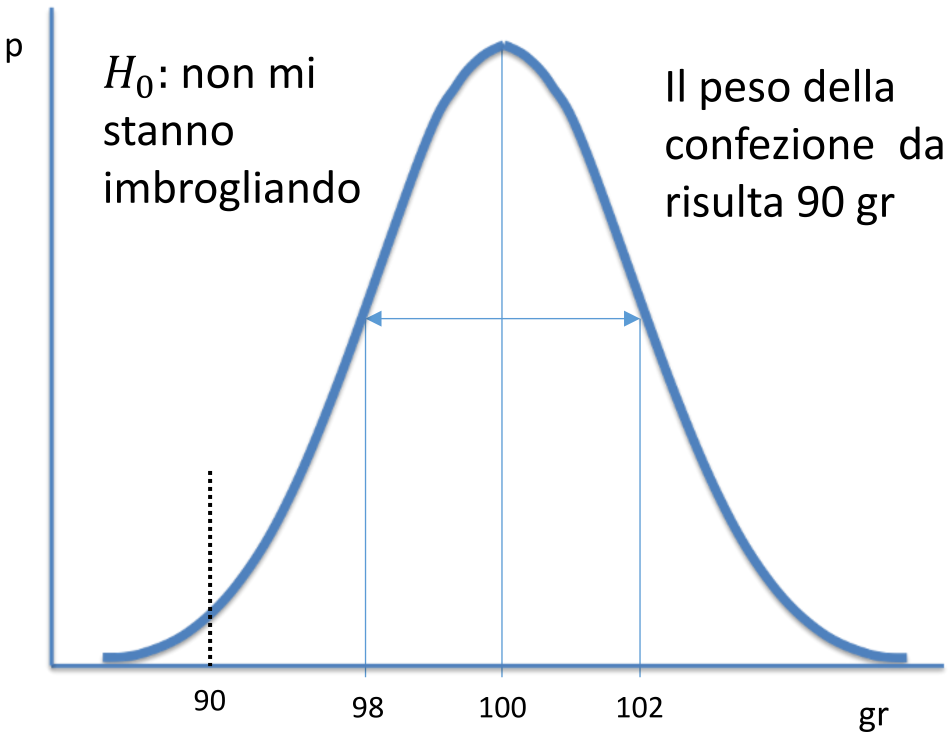


1. Si definisce l'ipotesi H_0. Il resto del metodo sarà svolto a dimostrare che H_0 è falsa
2. Si definisce la statistica che descrive il modello
3. Si osserva se e quanto la misura effettuata è improbabile data l'ipotesi H_0, calcolando la probabilità di ottenere un risultato uguale o ancora più lontano dall'ipotesi H_0 (p-value).

Il p-value e il test-ipotesi

Esempio: voglio assicurarmi di non essere imbrogliato sul peso della scatola dei biscotti, venduta per 100±2 gr

Il test-ipotesi è una procedura statistica per verificare con quale significatività una misura si discosta dal modello proposto.

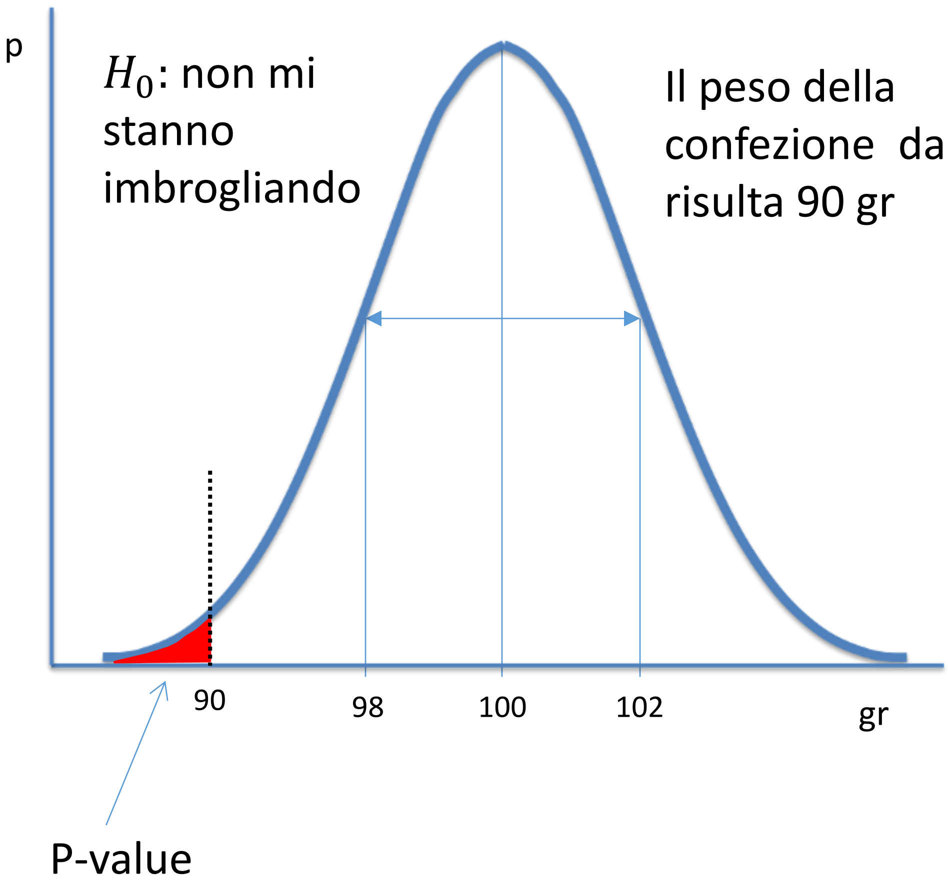


1. Si definisce l'ipotesi H_0. Il resto del metodo sarà svolto a dimostrare che H_0 è falsa
2. Si definisce la statistica che descrive il modello
3. Si osserva se e quanto la misura effettuata è improbabile data l'ipotesi H_0, calcolando la probabilità di ottenere un risultato uguale o ancora più lontano dall'ipotesi H_0 (p-value).

Il p-value e il test-ipotesi

Esempio: voglio assicurarmi di non essere imbrogliato sul peso della scatola dei biscotti, venduta per 100±2 gr

Il test-ipotesi è una procedura statistica per verificare con quale significatività una misura si discosta dal modello proposto.

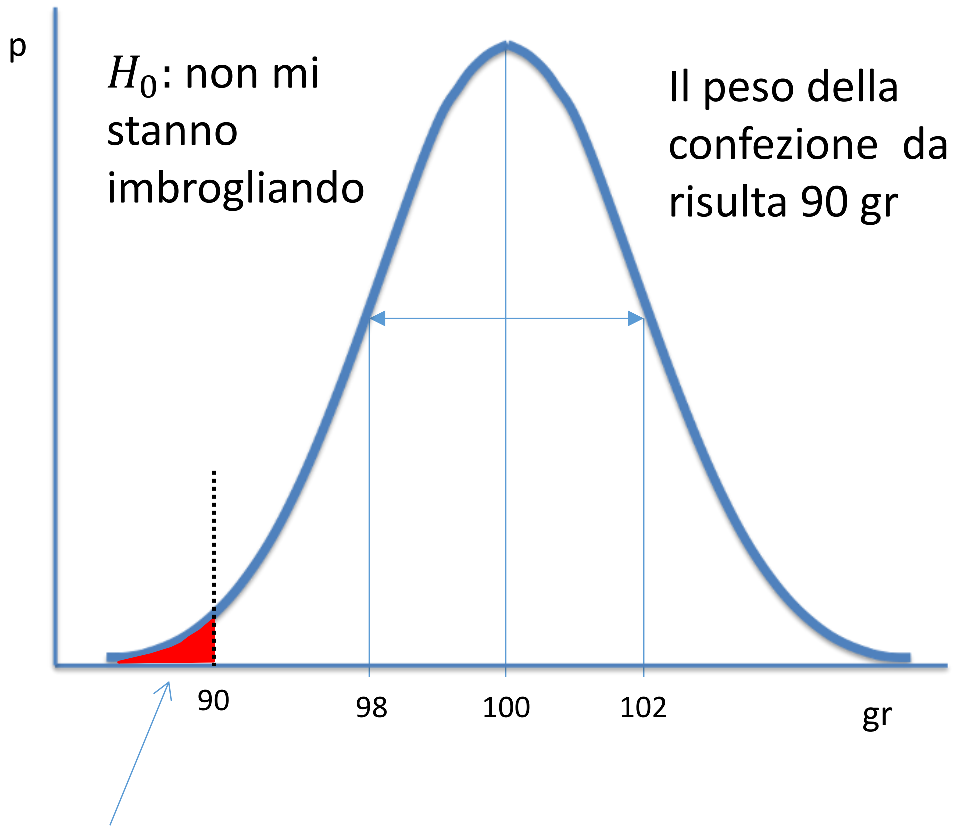


1. Si definisce l'ipotesi H_0. Il resto del metodo sarà svolto a dimostrare che H_0 è falsa
2. Si definisce la statistica che descrive il modello
3. Si osserva se e quanto la misura effettuata è improbabile data l'ipotesi H_0, calcolando la probabilità di ottenere un risultato uguale o ancora più lontano dall'ipotesi H_0 (p-value).

Il p-value e il test-ipotesi

Esempio: voglio assicurarmi di non essere imbrogliato sul peso della scatola dei biscotti, venduta per 100±2 gr

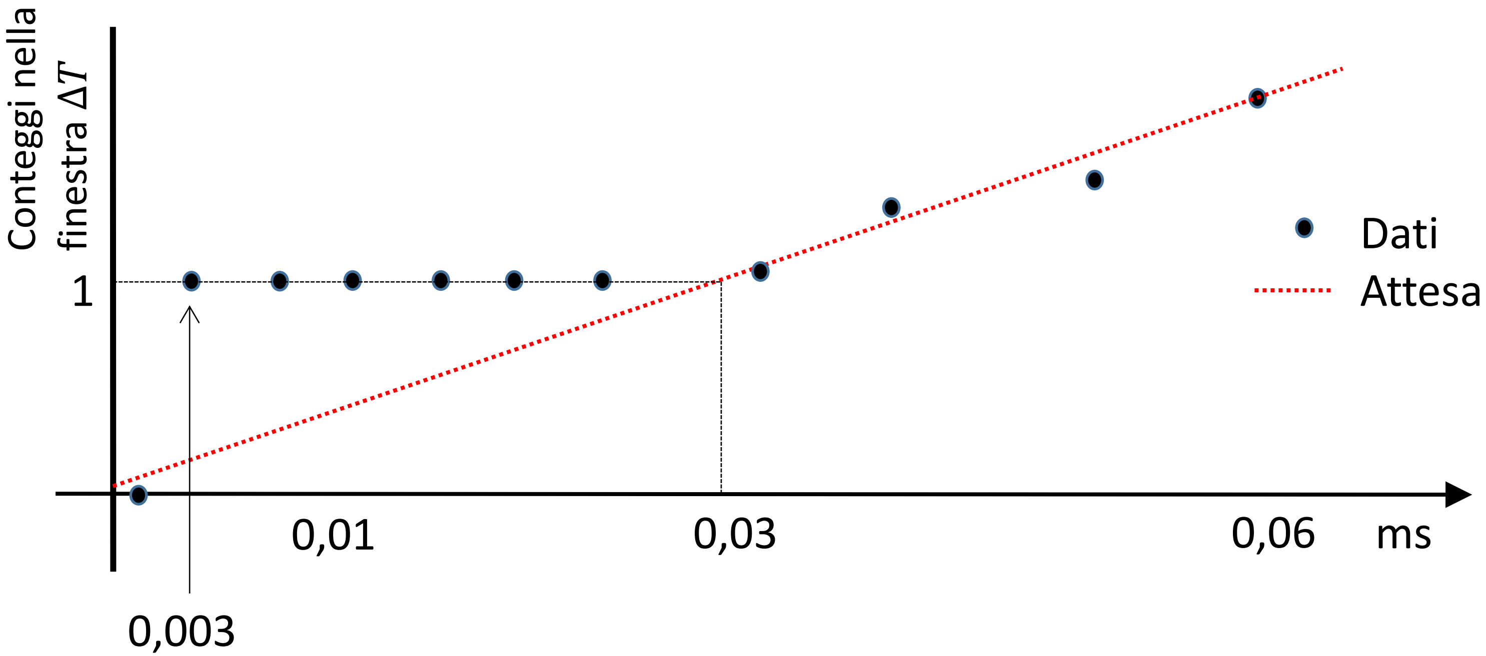
Il test-ipotesi è una procedura statistica per verificare con quale significatività una misura si discosta dal modello proposto.



1. Si definisce l'ipotesi H_0. Il resto del metodo sarà svolto a dimostrare che H_0 è falsa
2. Si definisce la statistica che descrive il modello
3. Si osserva se e quanto la misura effettuata è improbabile data l'ipotesi H_0, calcolando la probabilità di ottenere un risultato uguale o ancora più lontano dall'ipotesi H_0 (p-value).

P-value → 0,00006 % (forse devo cambiare marca…)

P-value per le LDC



Basso numero di conteggi all'interno di un intervallo temporale

Statistica di Poisson

$$P(k) = e^{-\mu} * \frac{\mu^k}{k!}$$

$\mu?$

H_0: non ci sono LDC --> I dati devono concordare con il modello lineare del fit.

Se H_0 è vera mi aspetto di trovare in media 1 coincidenza per finestre pari a 0,03 ms → μ=1

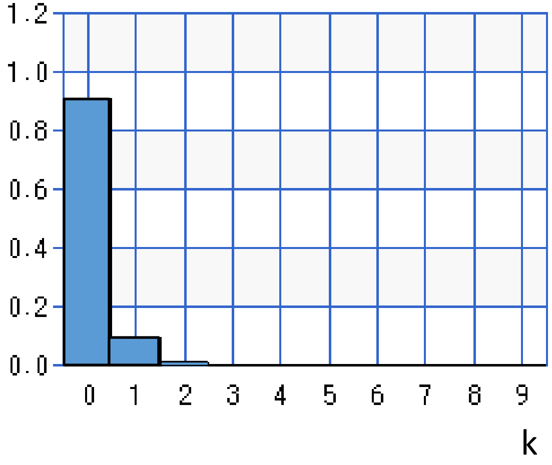
Quindi per finestre pari 0,003 ms (la più piccola in cui ho osservato un evento) → μ= 0,003/0,03= 0,1

P-value per le LDC

Quindi per finestre pari 0,003 ms (la più piccola in cui ho osservato un evento) →
μ= 0,003/0,03= 0,1

Statistica del modello con $\Delta T = 0{,}003$

$$P(k) = e^{-0{,}1} * \frac{0{,}1^k}{k!}$$



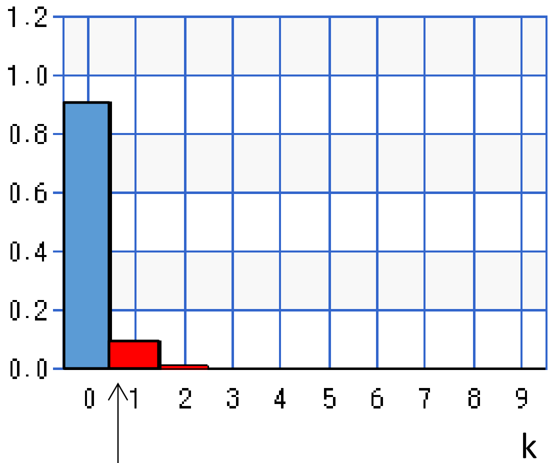
k	Prob.
0	0.904837
1	0.090484
2	0.004524
3	1.51E-04
4	3.77E-06
5	7.54E-08
6	1.26E-09
7	1.80E-11
8	2.24E-13
9	2.49E-15

P-value per le LDC

Quindi per finestre pari 0,003 ms (la più piccola in cui ho osservato un evento) →
μ= 0,003/0,03= 0,1

Statistica del modello con $\Delta T = 0{,}003$

$$P(k) = e^{-0{,}1} * \frac{0{,}1^k}{k!}$$



Abbiamo misurato un conteggio

k	Prob.
0	0.904837
1	0.090484
2	0.004524
3	1.51E-04
4	3.77E-06
5	7.54E-08
6	1.26E-09
7	1.80E-11
8	2.24E-13
9	2.49E-15

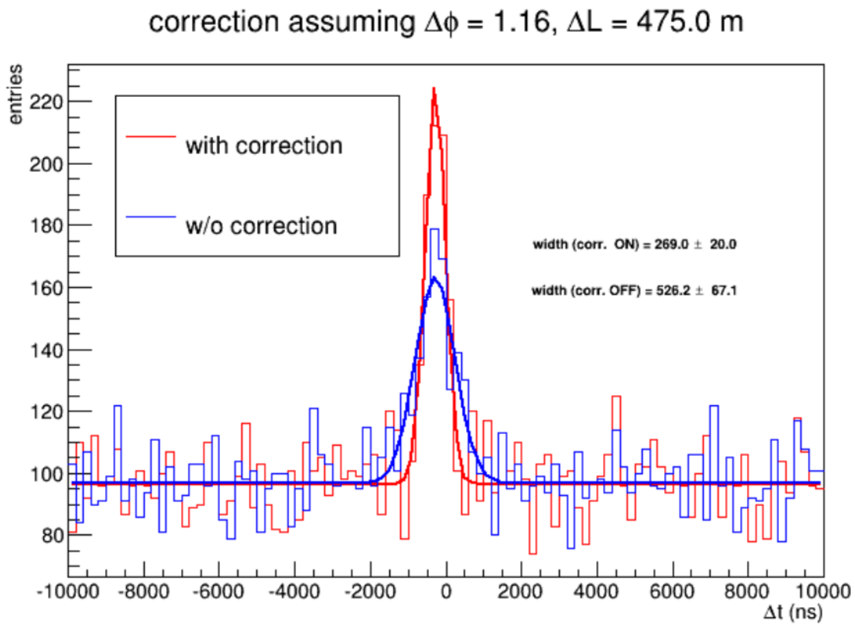
$$\begin{aligned} p-value &= P(\geq 1) \\ &= 1 - P(0) \\ &= 0{,}095 \end{aligned}$$

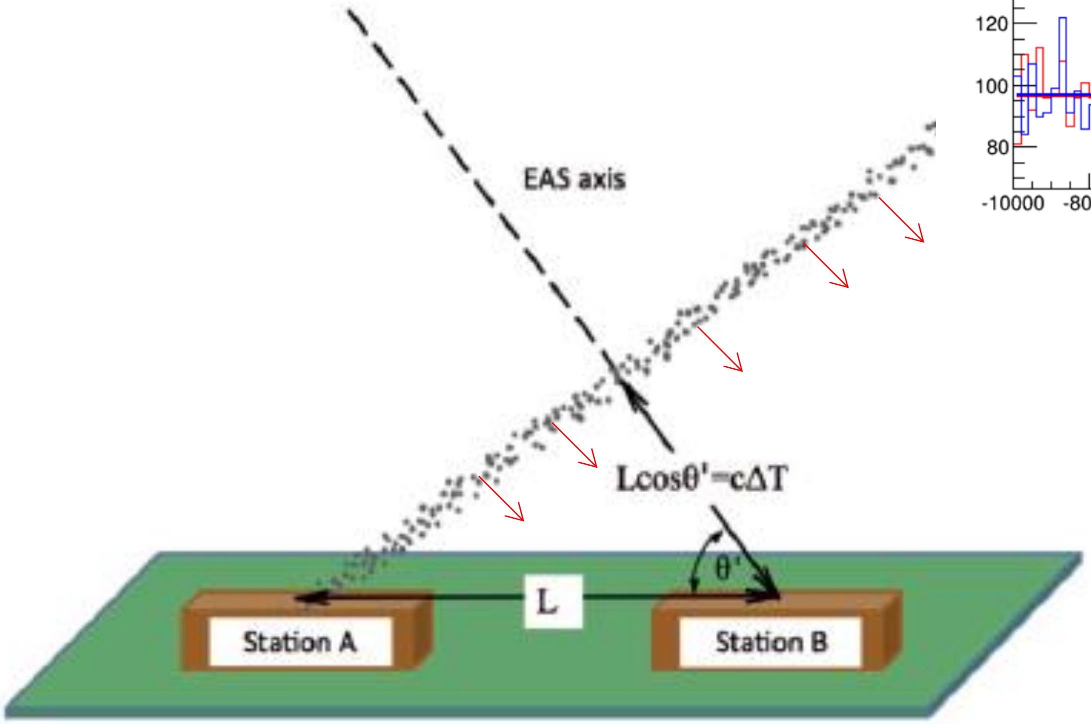
$3\sigma \cong 0{,}003$

$5\sigma \cong 0{,}0000003$

Esempio: osservazione di sciami estesi

Possiamo rinforzare i criteri per la ricerca delle coincidenze ed eliminare il «fondo» delle coincidenze accidentali





I muoni che riveliamo nei nostri telescopi sono prodotti a ~10 km di altezza. Due particelle generate dallo stesso primario e rivelate da due telescopi distanti ~500 metri saranno circa parallele.