

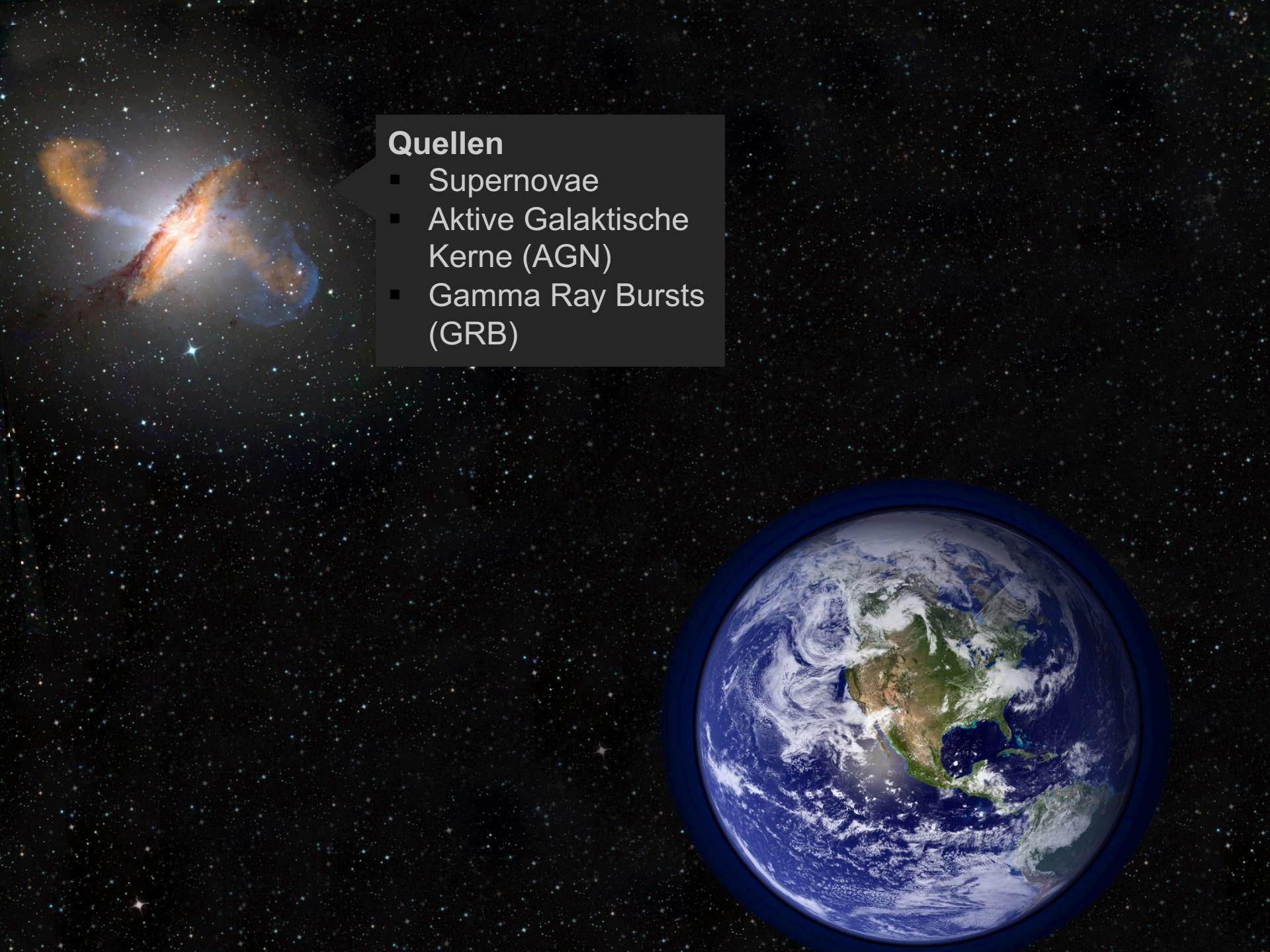
Vorlesung

Statistische Methoden der Datenanalyse

Prof. Dr. Dr. Wolfgang Rhode

Datenanalyse – Ein Beispiel aus der Praxis





Quellen

- Supernovae
- Aktive Galaktische Kerne (AGN)
- Gamma Ray Bursts (GRB)

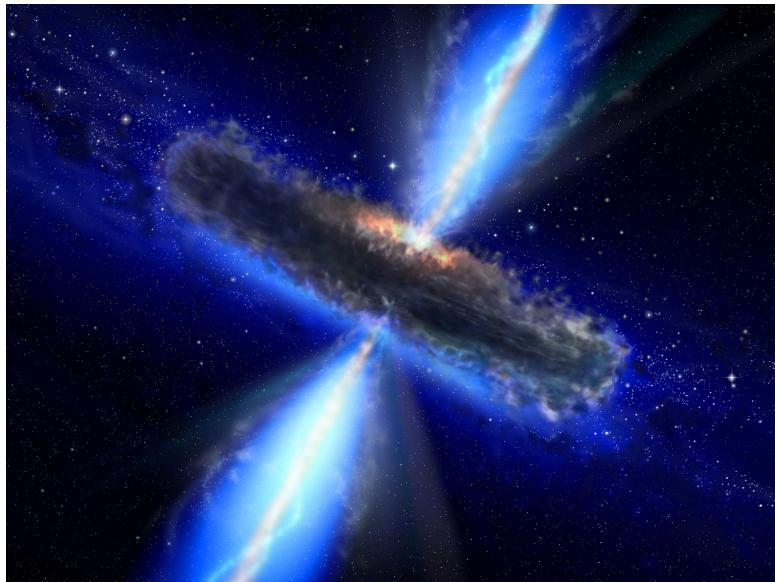


Supernovae



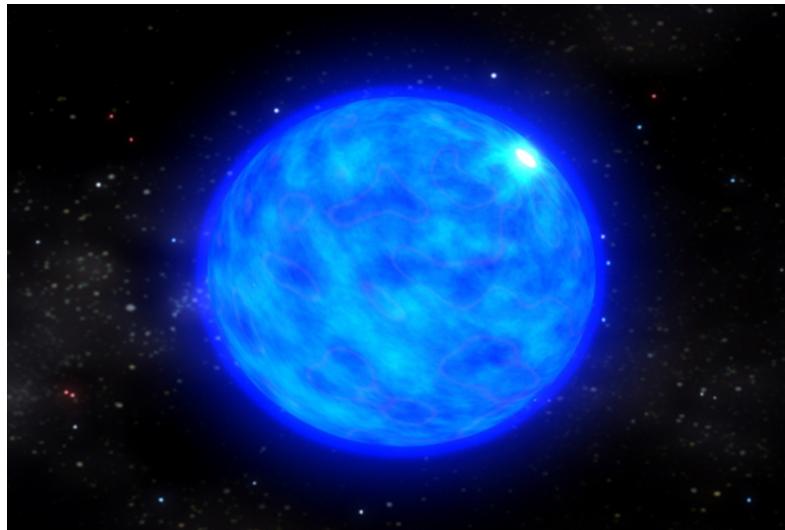
- Supernova
 - Ende Lebenszyklus eines Sternes; Fusionsmaterial *aufgebraucht*
 - Explosion mit hoher Emission von Photonen, geladenen Teilchen und Neutrinos
- Supernovaüberrest (SNR)
 - Expandierende Schockfront
 - Emittieren großen Teil der galaktischen kosmischen Strahlung
 - Beschleunigung durch den sogenannten Fermi Mechanismus

Aktive Galaktische Kerne (AGN)



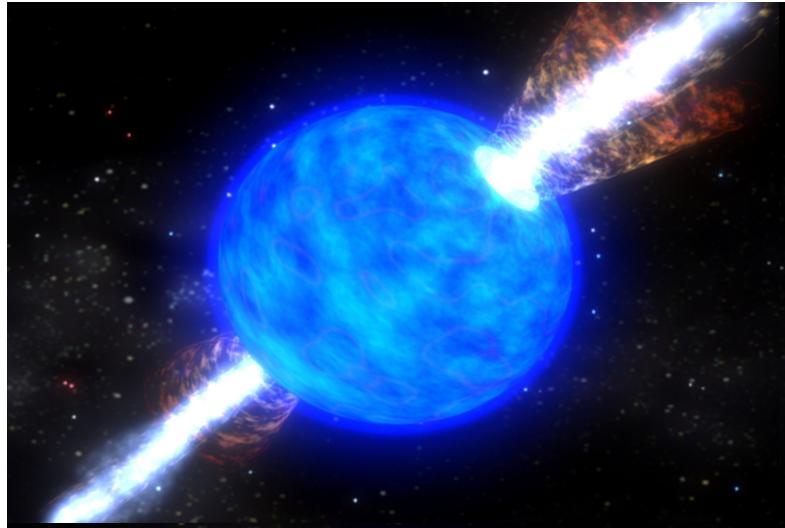
- Bestehen aus:
 - Supermassivem schwarzen Loch im Zentrum
 - Akkretionsscheibe um das Zentrum
 - Relativistischen Jets aus dem Zentrum, senkrecht zur Akkretionsscheibe
- Kategorisierung abhängig unserer Blickrichtung auf die Jets
- Emittieren hochenergetische kosmische Strahlung

Gamma Ray Bursts (GRB)



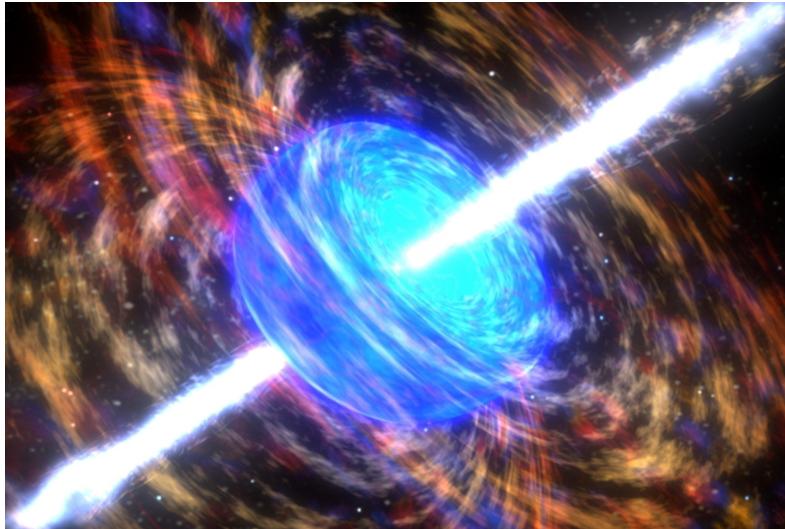
- Extrem hochenergetische Explosions in weit entfernten Galaxien
- Hohe Strahlungsemission während der Explosion
- Dauer in der Größenordnung von Sekunden

Gamma Ray Bursts (GRB)



- Extrem hochenergetische Explosions in weit entfernten Galaxien
- Hohe Strahlungsemission während der Explosion
- Dauer in der Größenordnung von Sekunden

Gamma Ray Bursts (GRB)



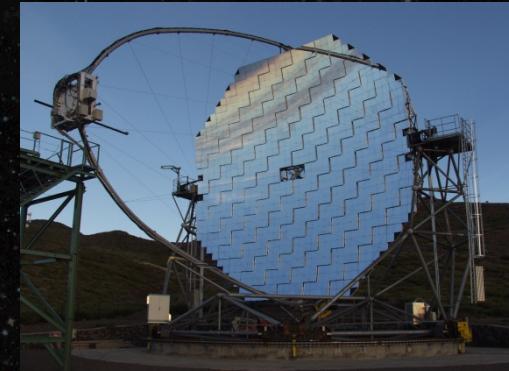
- Extrem hochenergetische Explosions in weit entfernten Galaxien
- Hohe Strahlungsemission während der Explosion
- Dauer in der Größenordnung von Sekunden



- 
- Unabgelenkt
 - Absorption
 - Meist aus den äußeren Bereichen der Quellen

Photonen

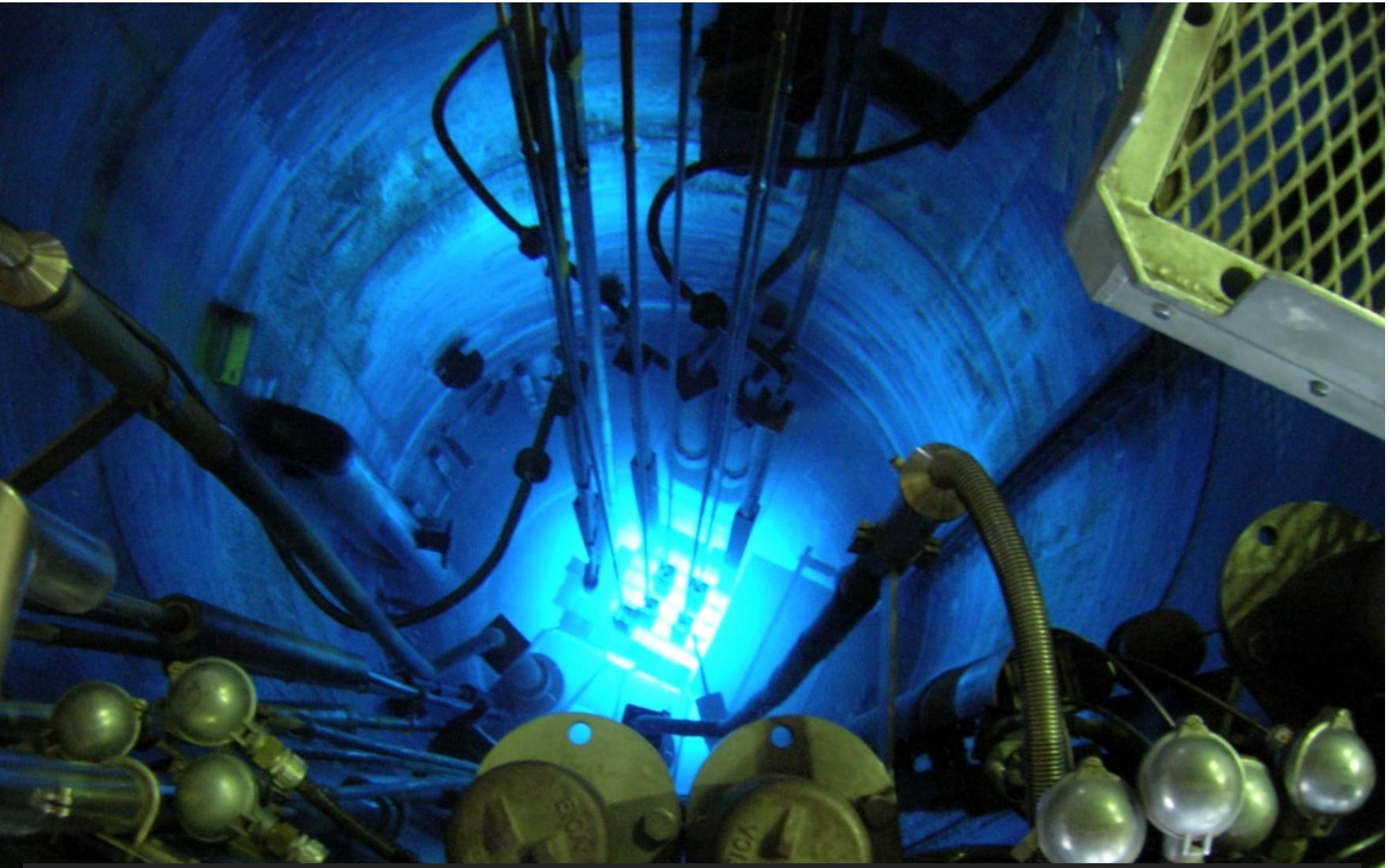
- Optisch
- Röntgen
- Mikrowellen
- Gamma



Photonen



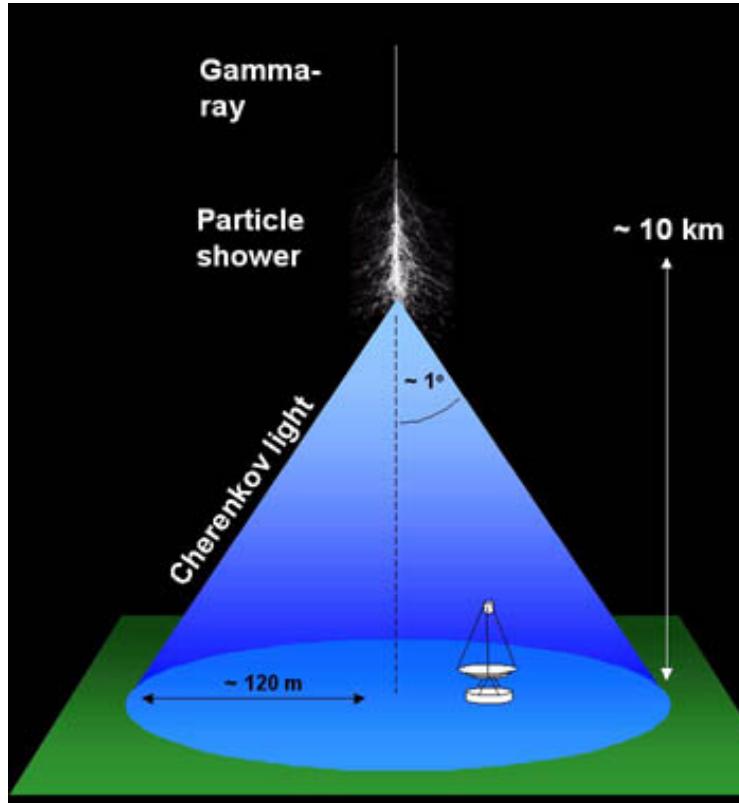
- Direkter Nachweis/Messung der Photonen
 - Sichtbares Licht
 - Radiowellen
 - Röntgen (mit Satelliten)



Tscherenkov Strahlung:

Geladene Teilchen, die sich durch ein Medium bewegen und dabei schneller sind als die Lichtgeschwindigkeit in dem Medium, emittieren Photonen.

Photonen



- Indirekter Nachweis/Messung hochenergetischer Gamma Strahlung
 - Messung der Tscherenkov Photonen, die vom elektromagnetischen Teilchenschauer emittiert werden

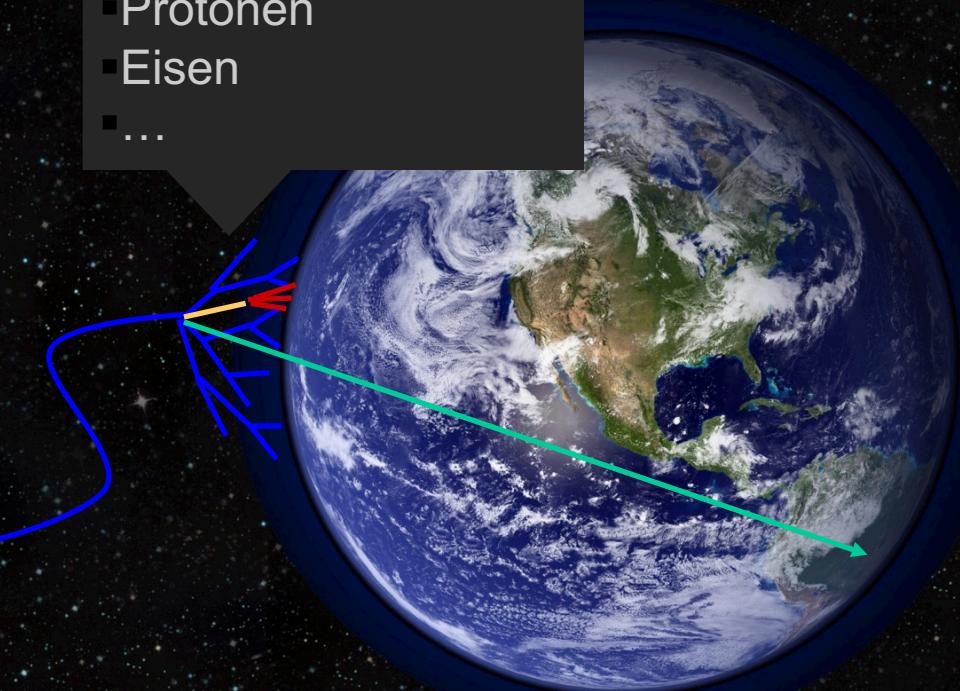
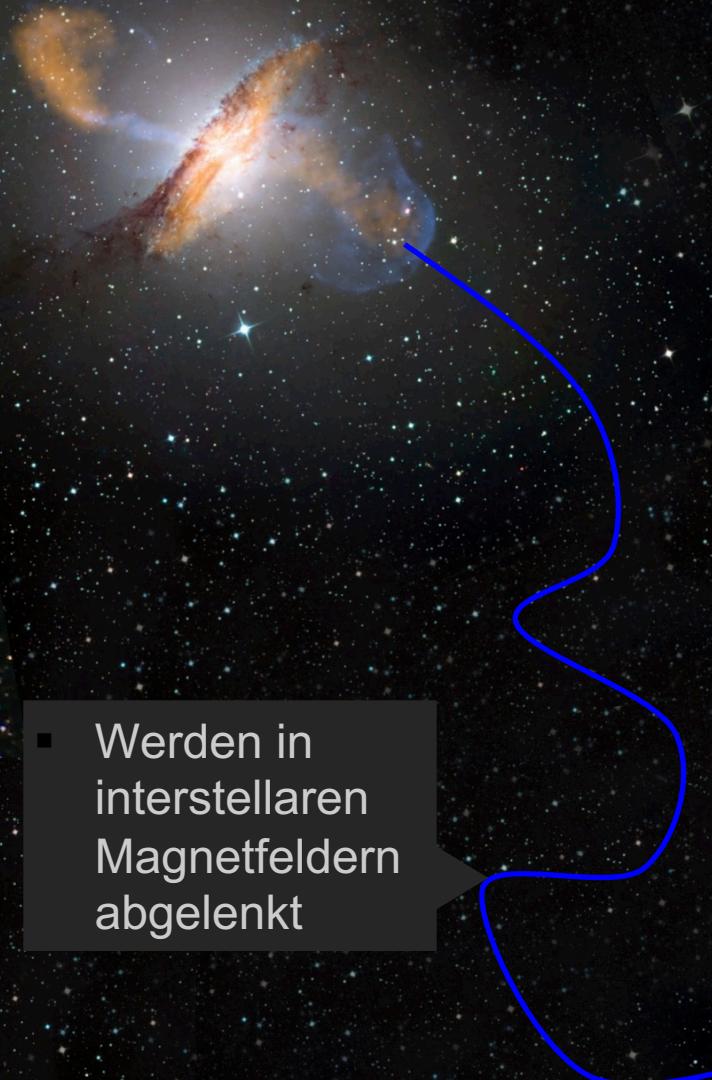
Photonen



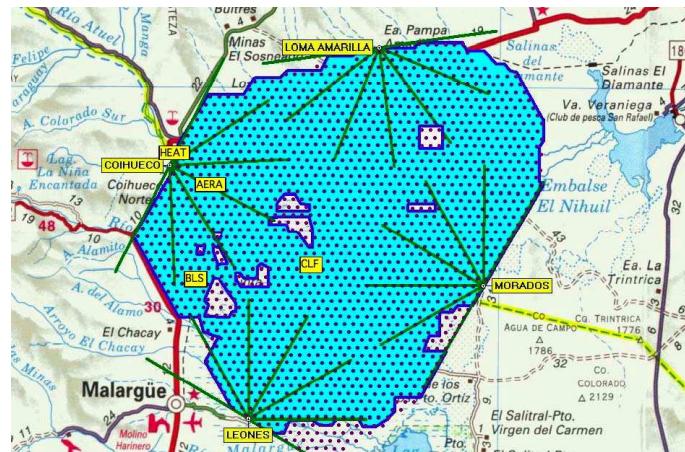
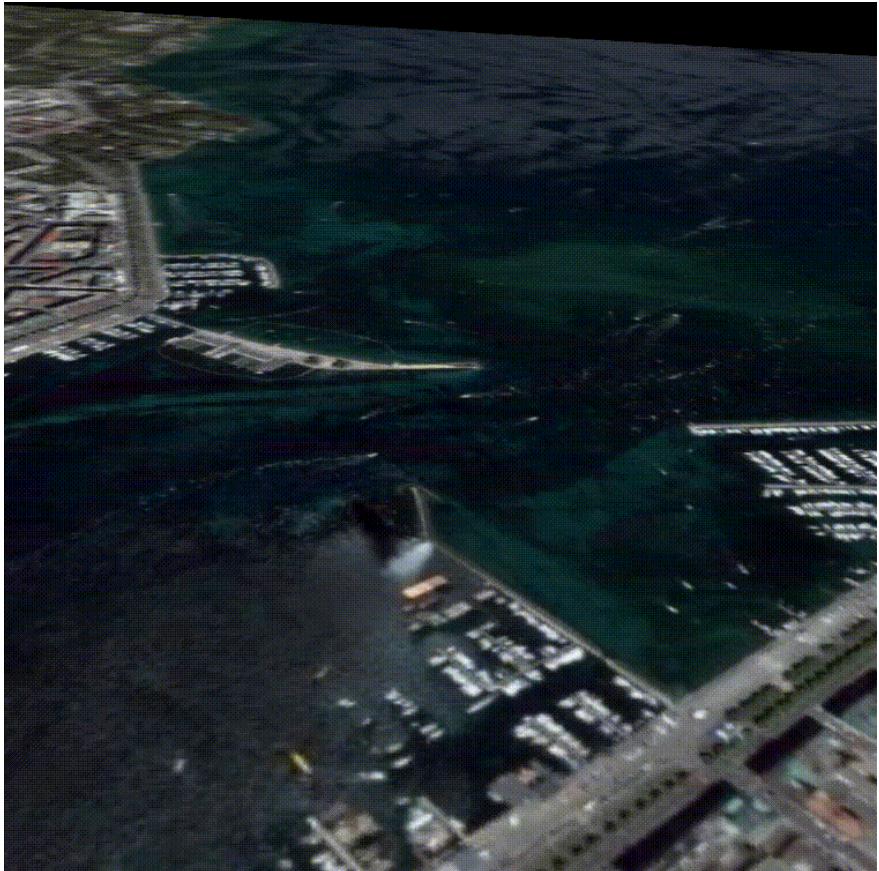
- Direkter Nachweis/Messung der Photonen
 - Sichtbares Licht
 - Radiowellen
 - Röntgen (mit Satelliten)
- Indirekter Nachweis/Messung hochenergetischer Gamma Strahlung
 - Messung der Tscherenkov Photonen, die vom elektromagnetischen Teilchenschauer emittiert werden



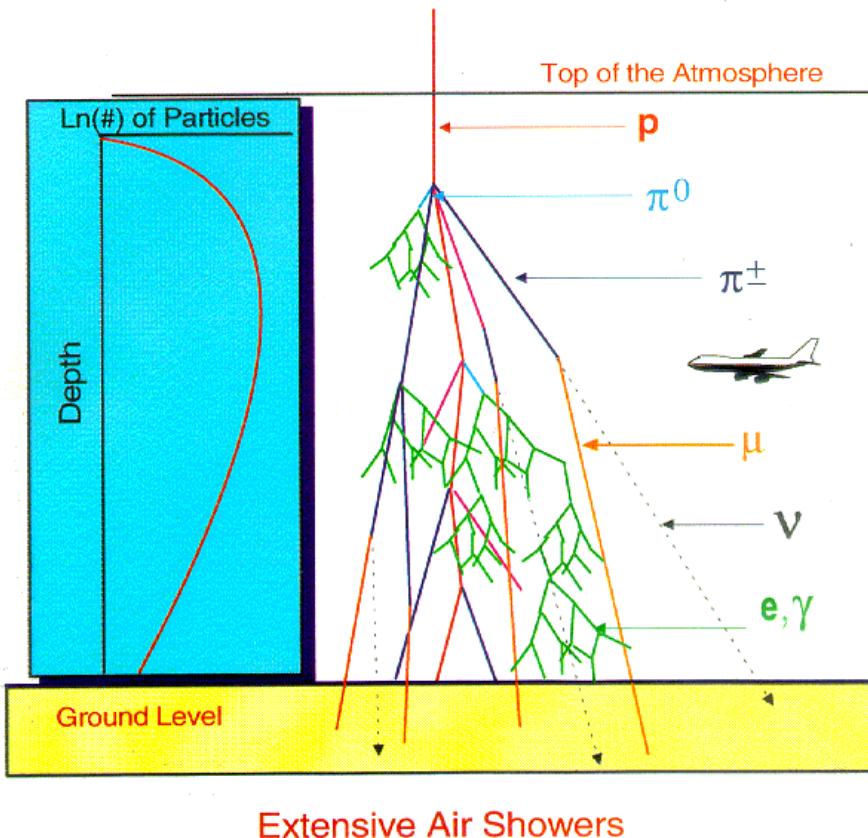
- Werden in interstellaren Magnetfeldern abgelenkt



Geladene Strahlung

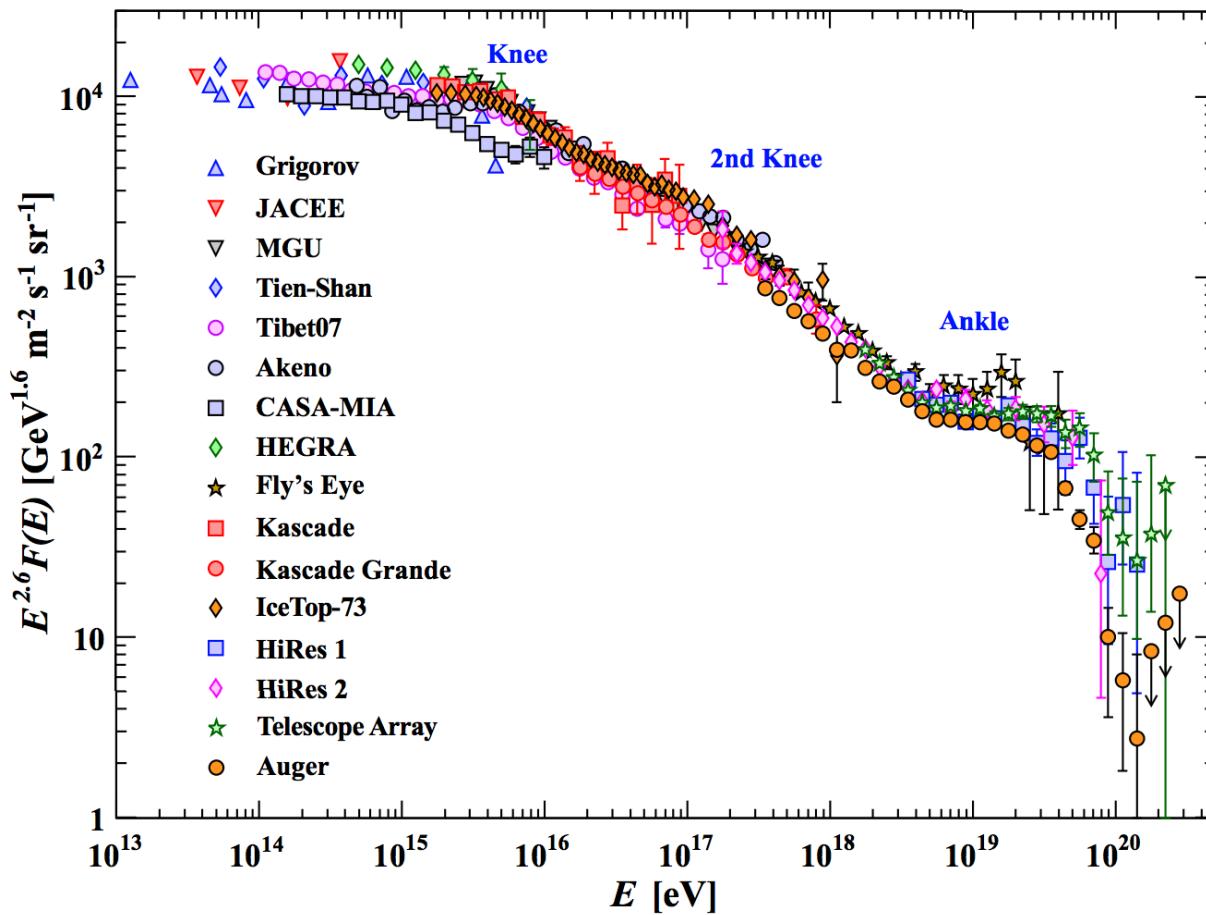


Geladene Strahlung



- Trifft ein hochenergetisches Teilchen der geladenen kosmischen Strahlung auf der Atmosphäre entsteht ein Luftschauder
- Teilchen, die in diesen Schauern entstehen, können direkt gemessen werden
- Indirekte Messung:
 - Tscherenkov Strahlung
 - Fluoreszenzlicht von in der Luft angeregten Stickstoffmolekülen

Geladene Strahlung



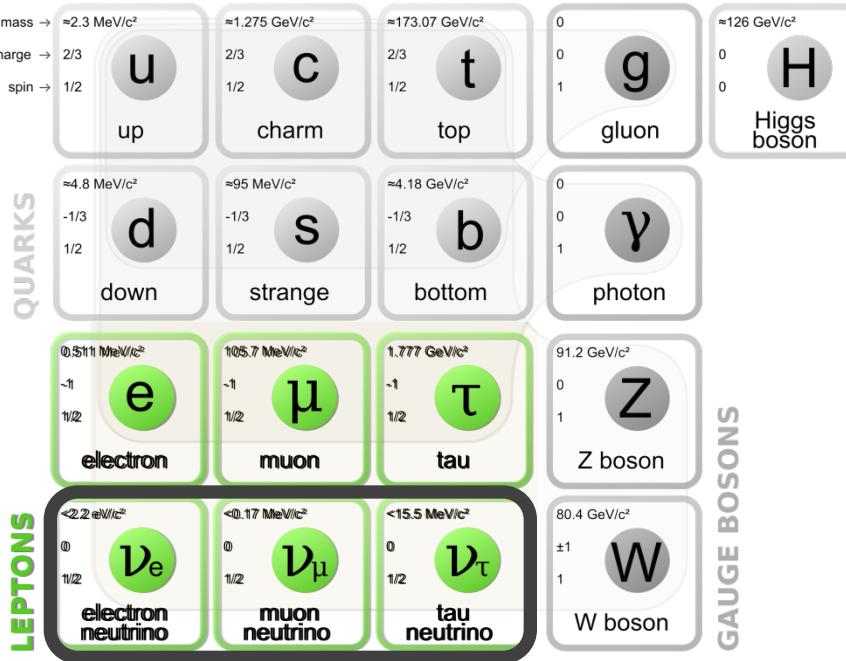




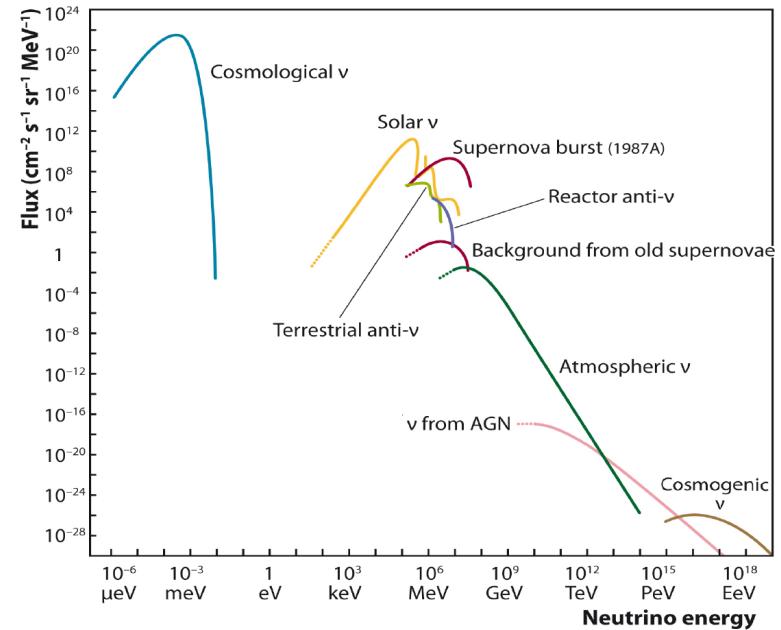
Neutrinos

- Solar
- Supernova
- Kosmisch

Neutrinos



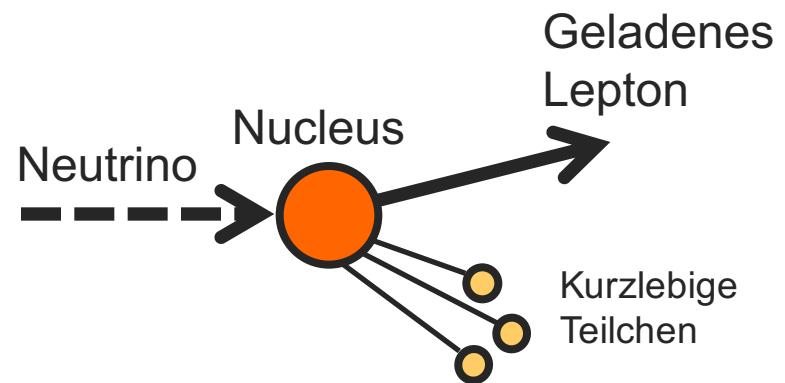
- Neutrinos entstehen in hadronischen Wechselwirkungen



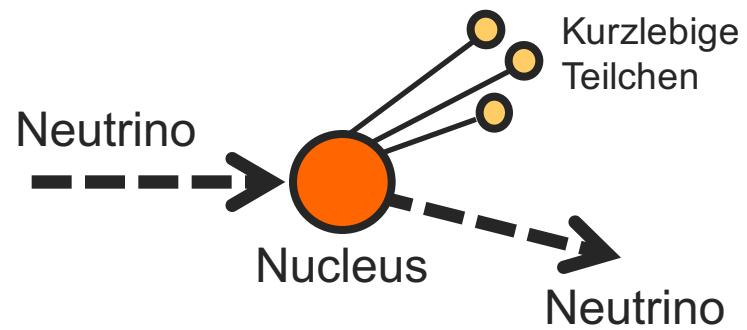
Neutrinos

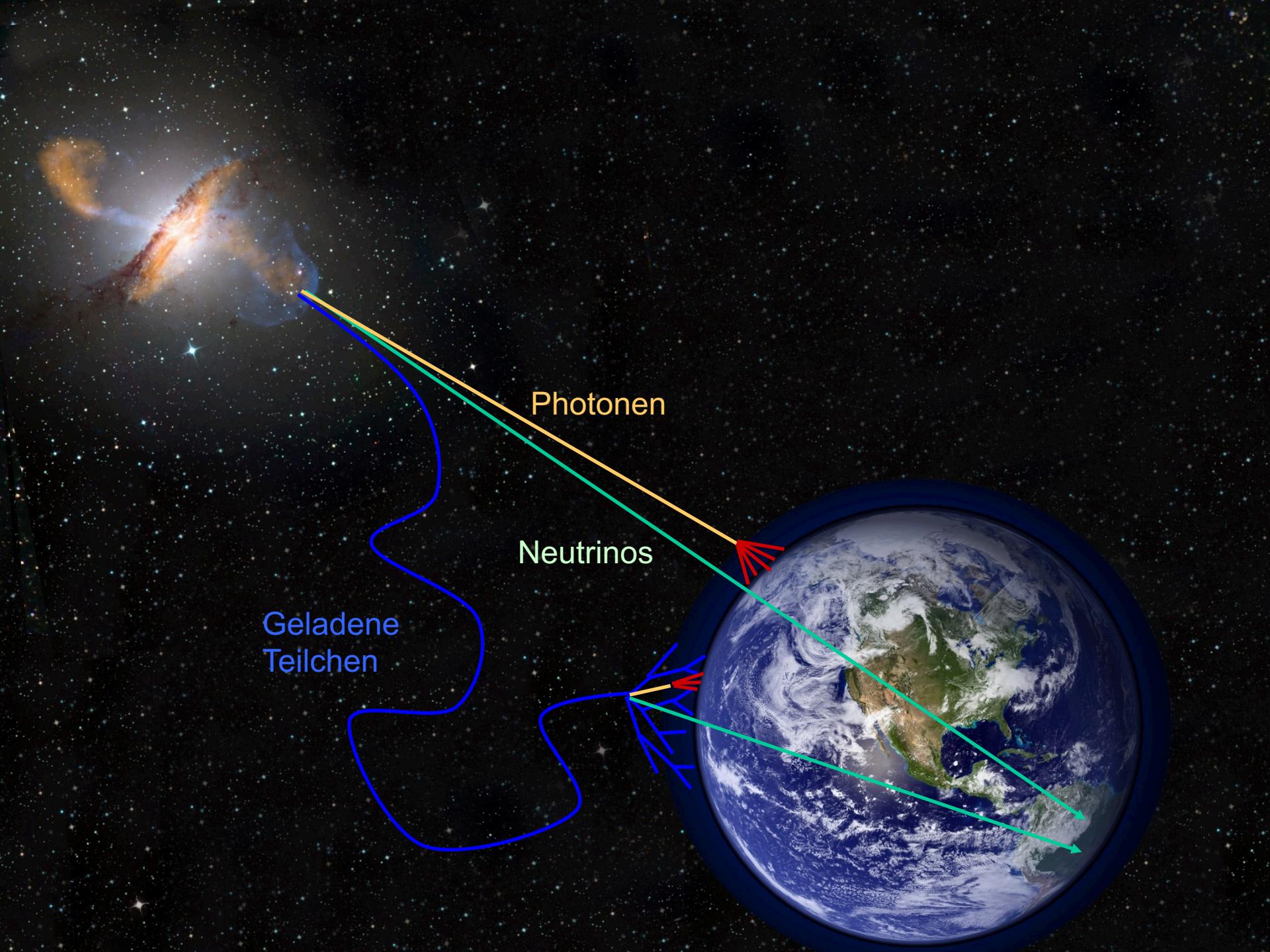
- Da Neutrinos nur Teil der schwachen Wechselwirkung sind, ist eine direkte Messung nicht möglich
 - Messung der Tscherenkov Photonen, die von Teilchen, die nach der Neutrinowechselwirkung entstehen, emittiert werden
- Quellen für hochenergetische kosmische Neutrinos ($> 100 \text{ TeV}$) unbekannt
- Experimenteller Nachweis ihrer Existenz *erst Ende 2013* durch IceCube

CC-Wechselwirkung:



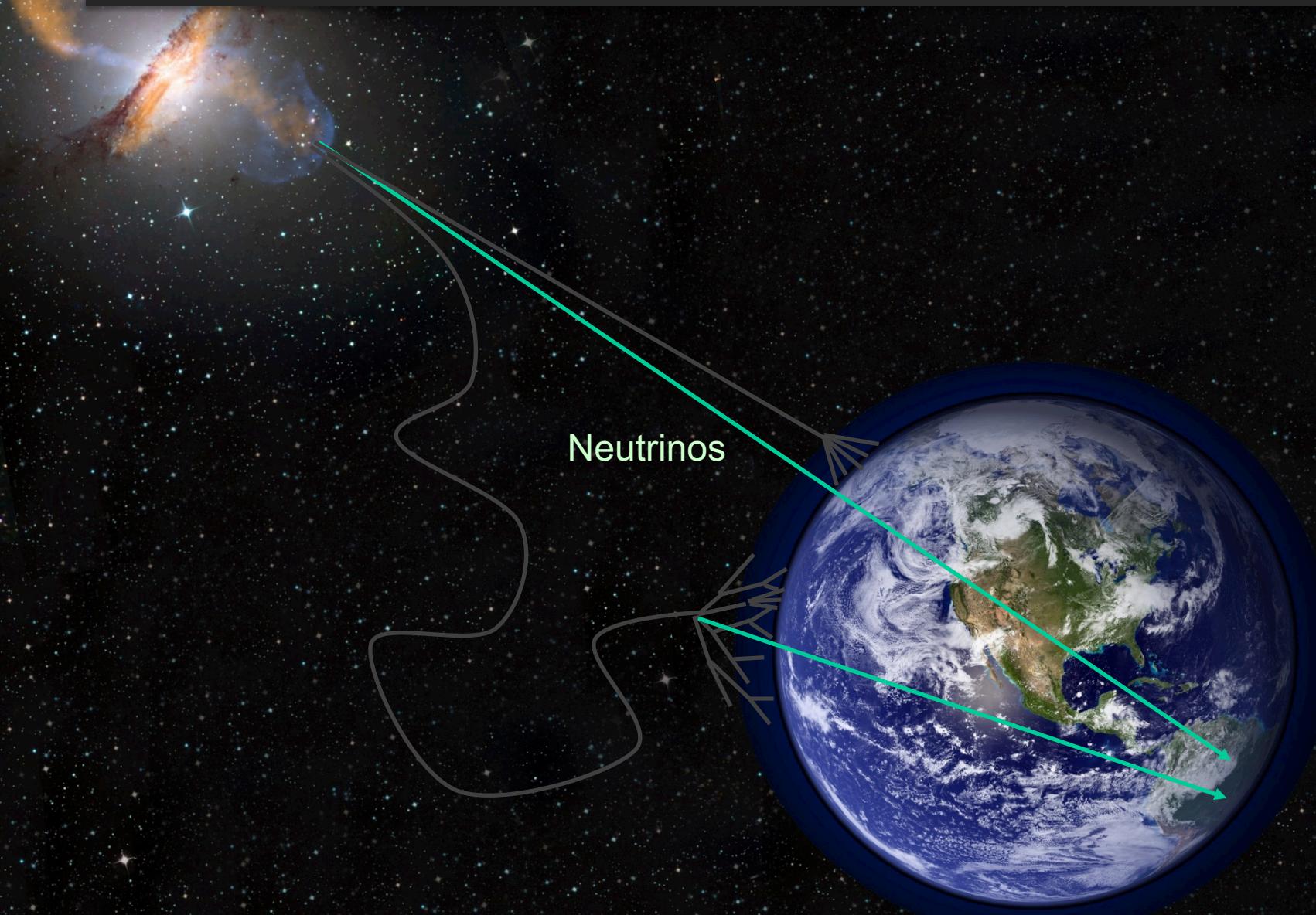
NC-Wechselwirkung:





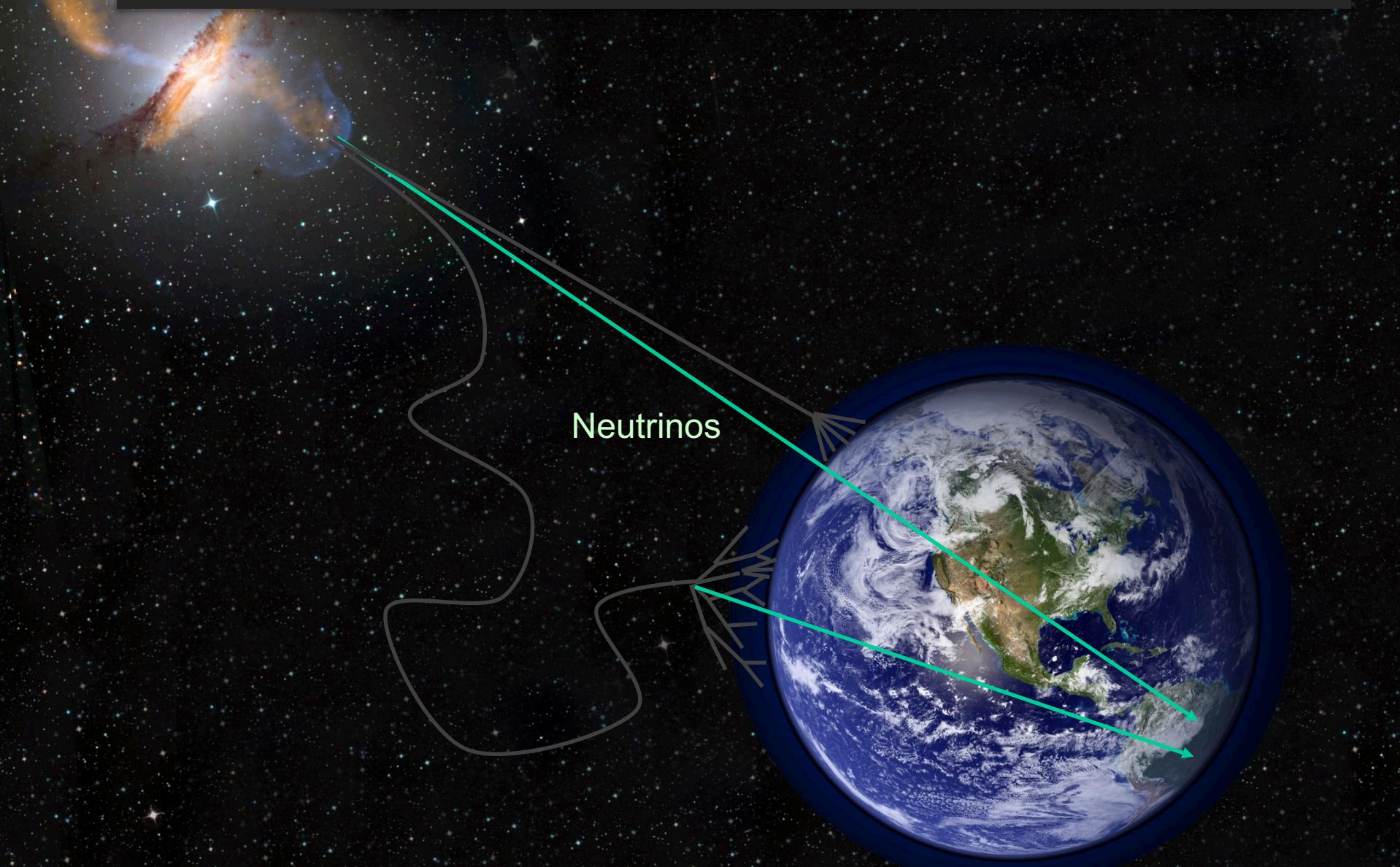
Datenanalyse am Beispiel:

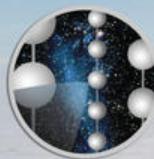
Messung des Energiespektrums von Myon-Neutrinos in IceCube



Datenanalyse am Beispiel:

Messung des Energiespektrums von Myon-Neutrinos in IceCube





ICECUBE

SOUTH POLE NEUTRINO OBSERVATORY

50 m



IceCube Laboratory

Data is collected here and sent by satellite to the data warehouse at UW–Madison

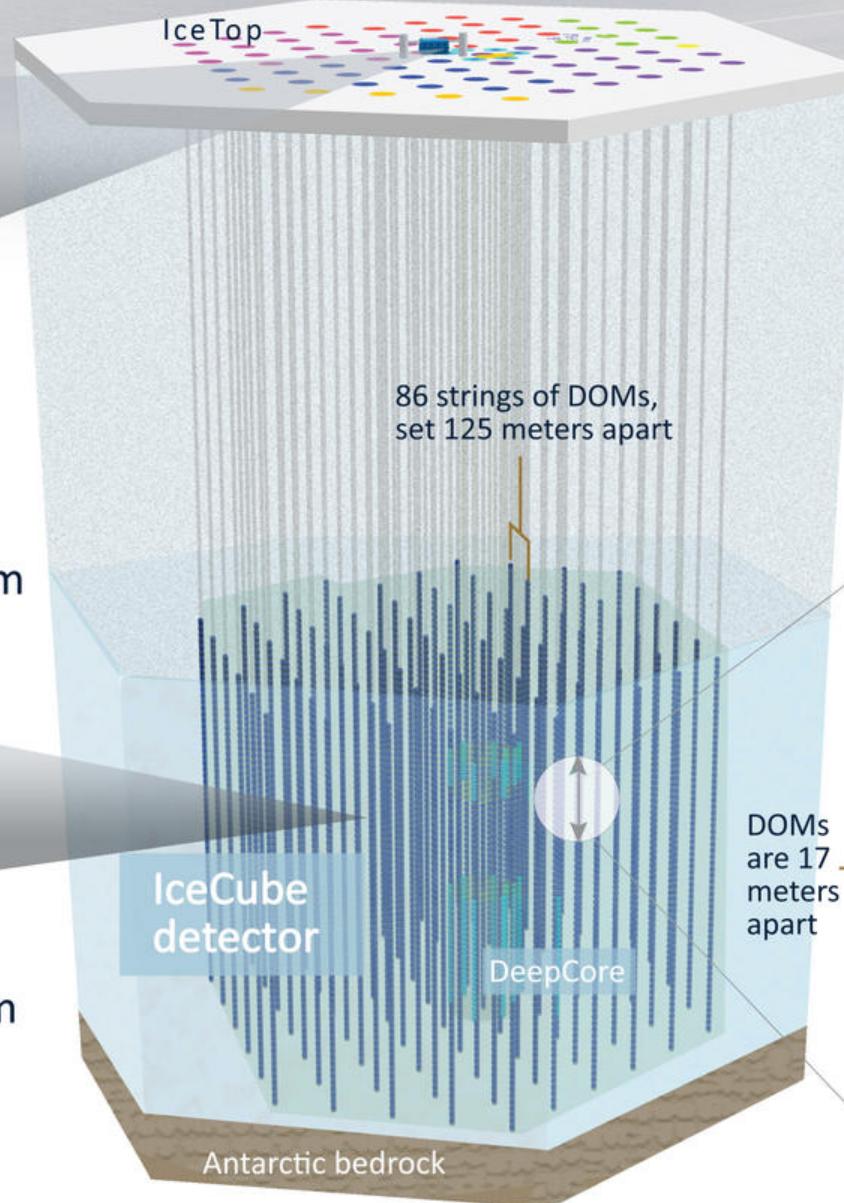
1450 m



Digital Optical Module (DOM)

5,160 DOMs deployed in the ice

2450 m



Amundsen–Scott South Pole Station, Antarctica
A National Science Foundation-managed research facility

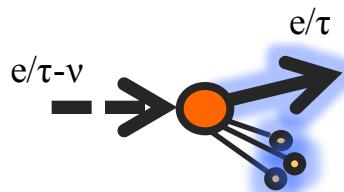
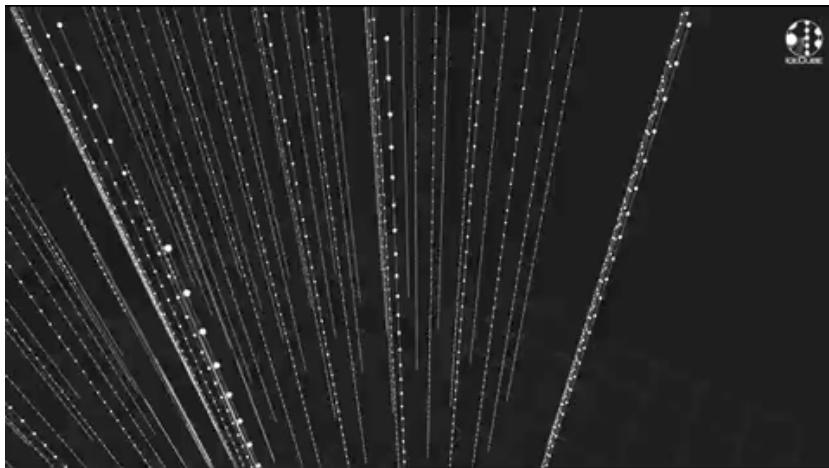
60 DOMs on each string

DOMs are 17 meters apart

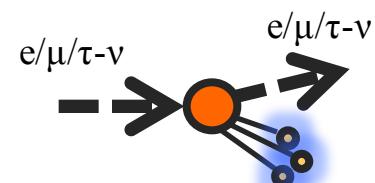


IceCube – Event Signaturen

Kaskaden

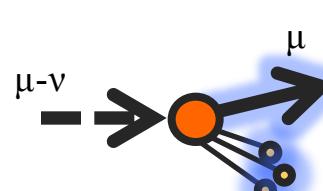
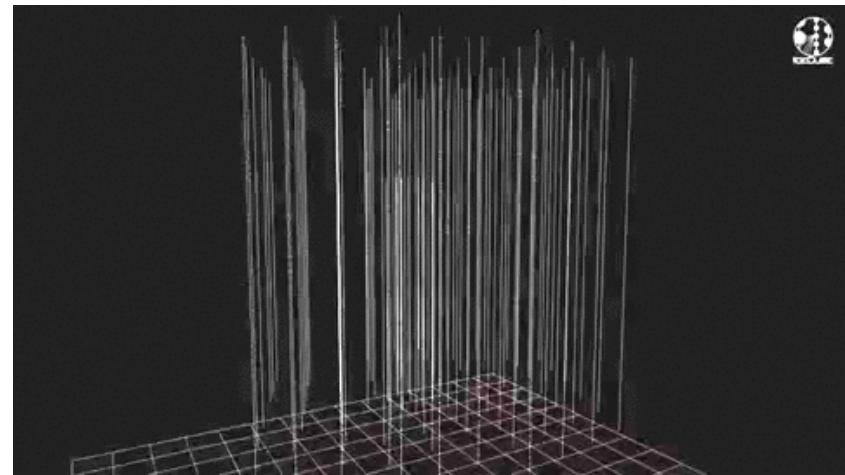


CC-Wechselwirkung
 e/τ -Neutrino

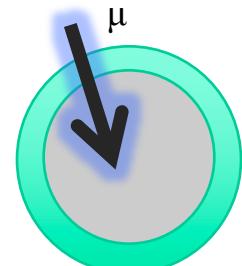


NC-Wechselwirkung
 $e/\mu/\tau$ -Neutrino

Spurartige Ereignisse (Tracks)



CC-Wechselwirkung
 μ -Neutrino

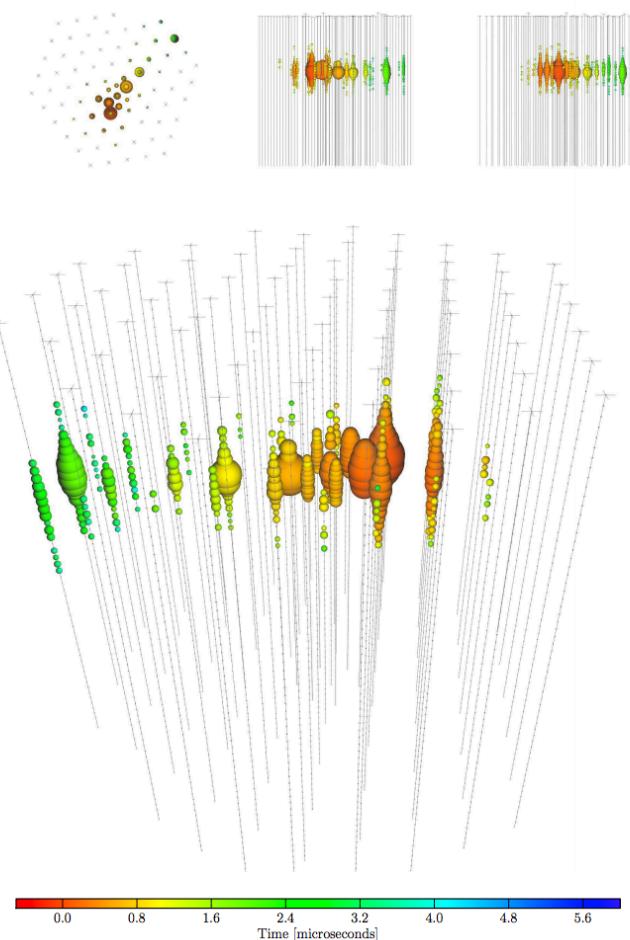


atmosphärisches μ
Statistische Methoden
der Datenanalyse

IceCube - Rekonstruktionen

- Die tatsächliche Messung eines Ereignisses besteht aus über fünftausend 190ns langen Spannungszeitreihen (für jeden DOM eine)
- Rekonstruktionen von physikalischen Observablen nötig

Splitting/Cleaning
<ul style="list-style-type: none"> ■ Koinzidenzen ■ Rauschen
Richtungsrekonstruktion
<ul style="list-style-type: none"> ■ Kaskaden → ungenau ■ Tracks → genau
Energierekonstruktion
<ul style="list-style-type: none"> ■ Kaskaden → genau ■ Tracks → ungenau



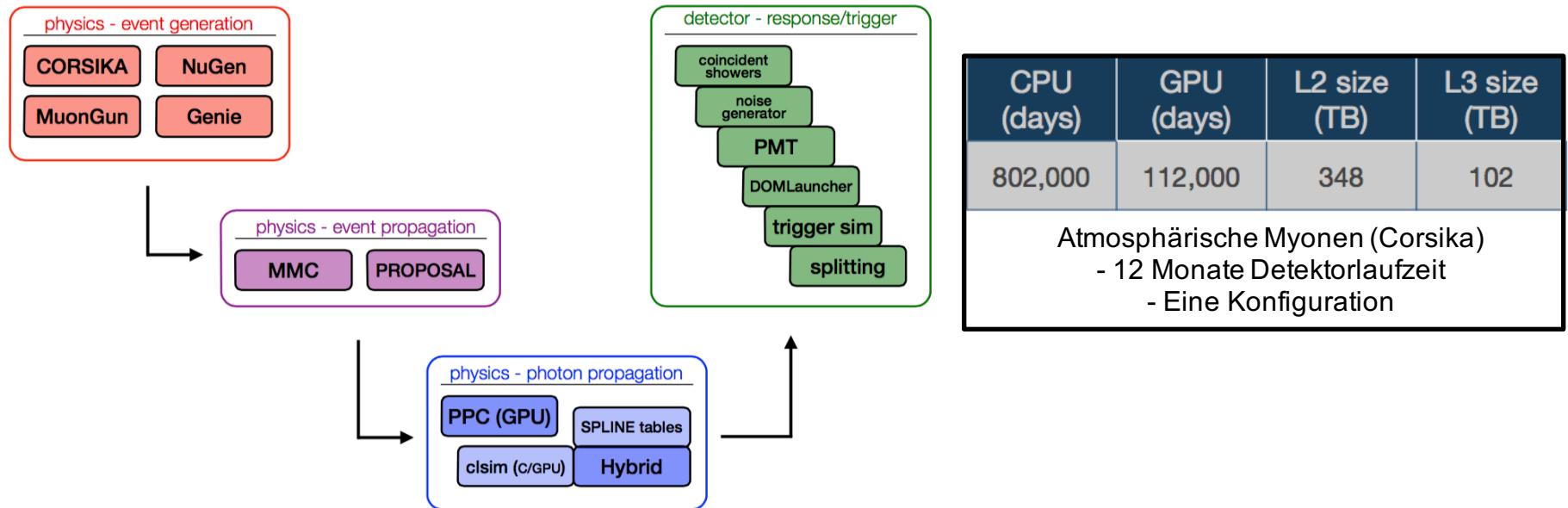
Deposited Energy (TeV)	Time (MJD)	Declination (deg.)	RA (deg.)	Med. Ang. Resolution (deg.)	Topology
71.4 ^{+9.0} _{-9.0}	55512.5516214	-0.4	110.6	$\lesssim 1.2$	Track

„Starting track“-Event:

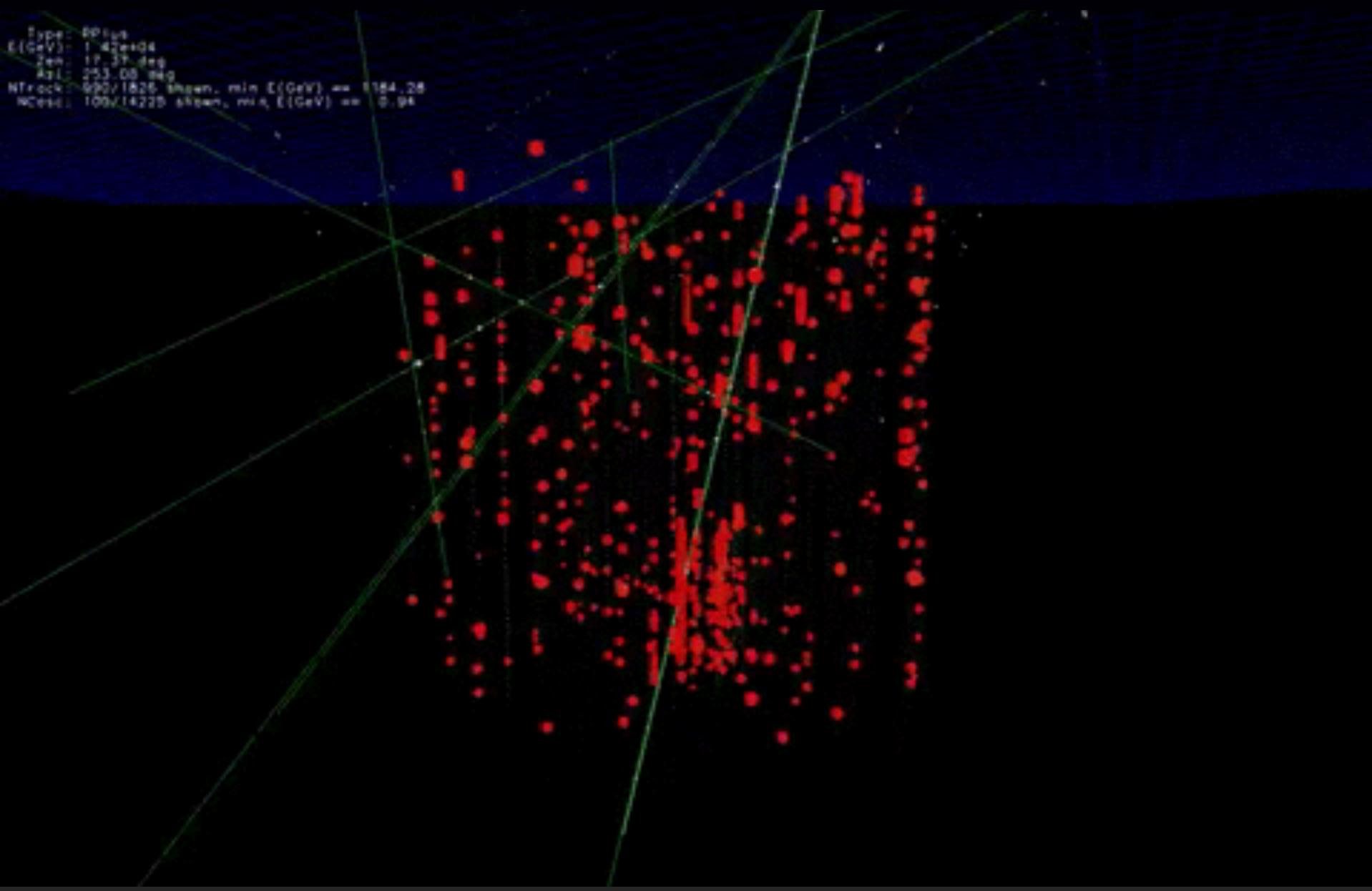
- Größe der Kugeln proportional zur integrierten Spannung
- Farbe assoziiert mit Zeitpunkt der Ladungsdeposition

IceCube - Simulationen

- Kalibration mit *Test-Neutrinos* unmöglich und viele moderne Analyseverfahren, die auf Simulationen basieren, im Einsatz
- Extrem aufwändige und detaillierte Simulationen startend bei der Teilchenproduktion bis zum Ende der Detektion nötig

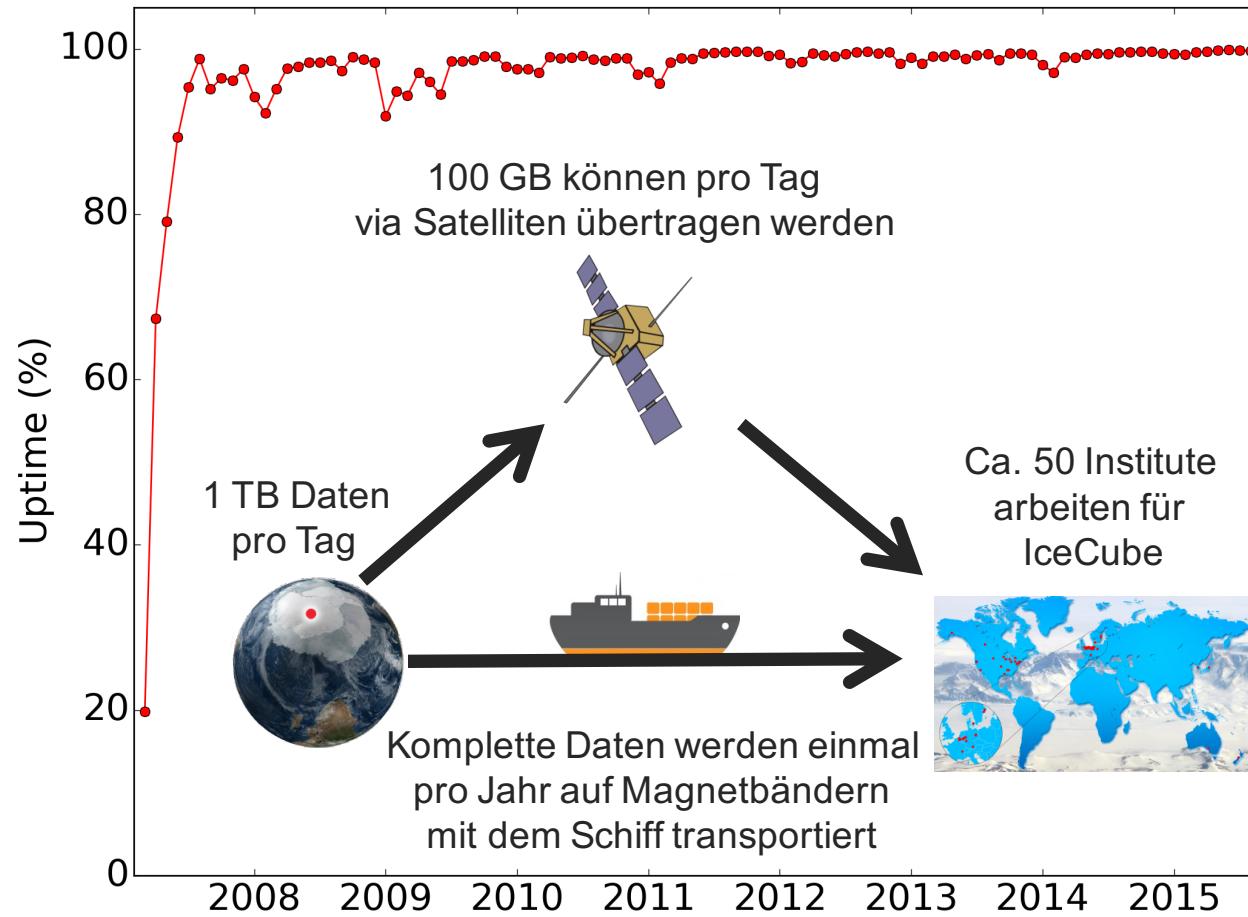


Type: DPhi vs.
E(GeV)= 1.42e+04
Zen: 17.37 deg
AzL: 253.08 deg
NTrock: 990/1826 shown, min E(GeV) == 1184.26
NCess: 100/14225 shown, min E(GeV) == 0.98



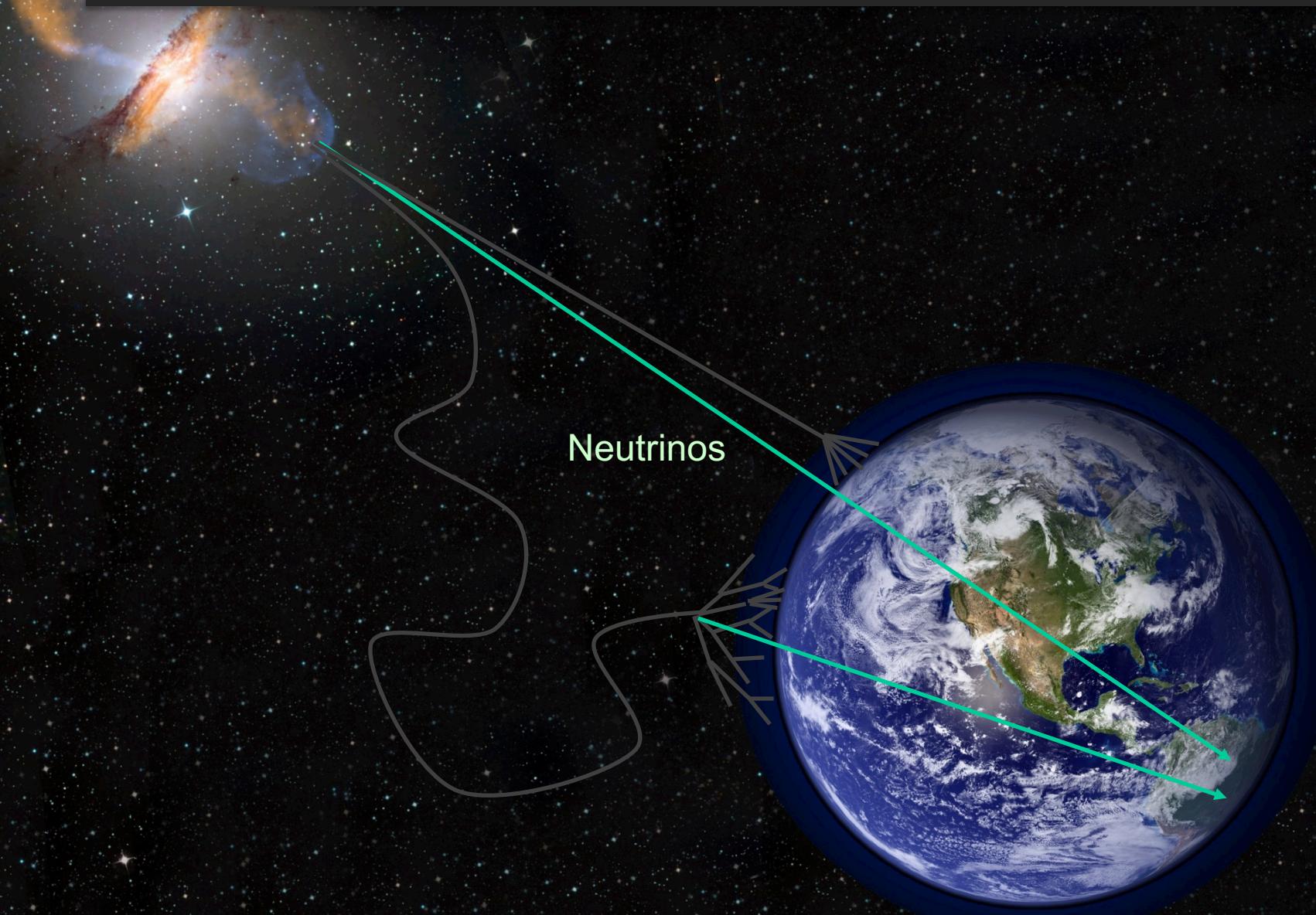
10ms in IceCube

IceCube – Big Data



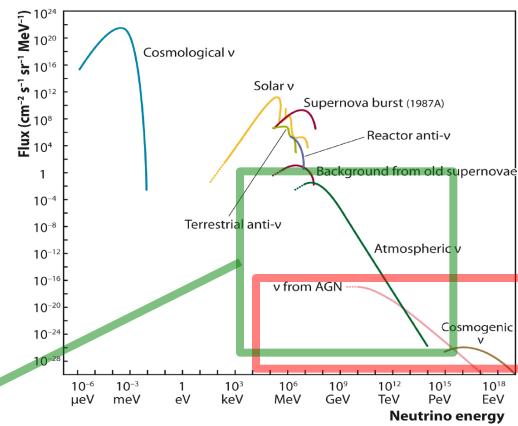
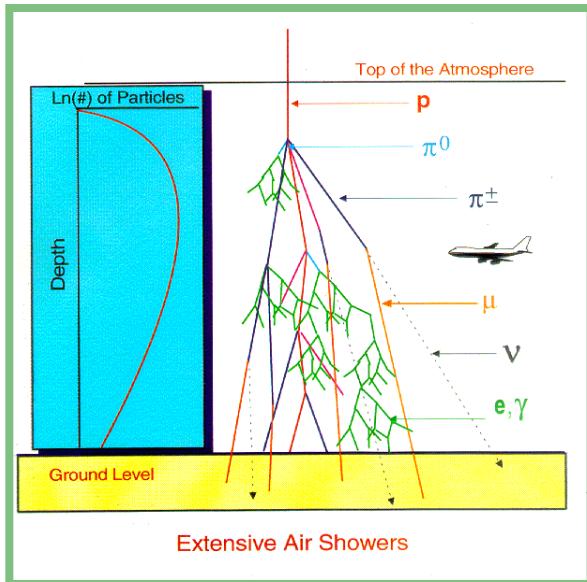
Datenanalyse am Beispiel:

Messung des Energiespektrums von Myon-Neutrinos in IceCube



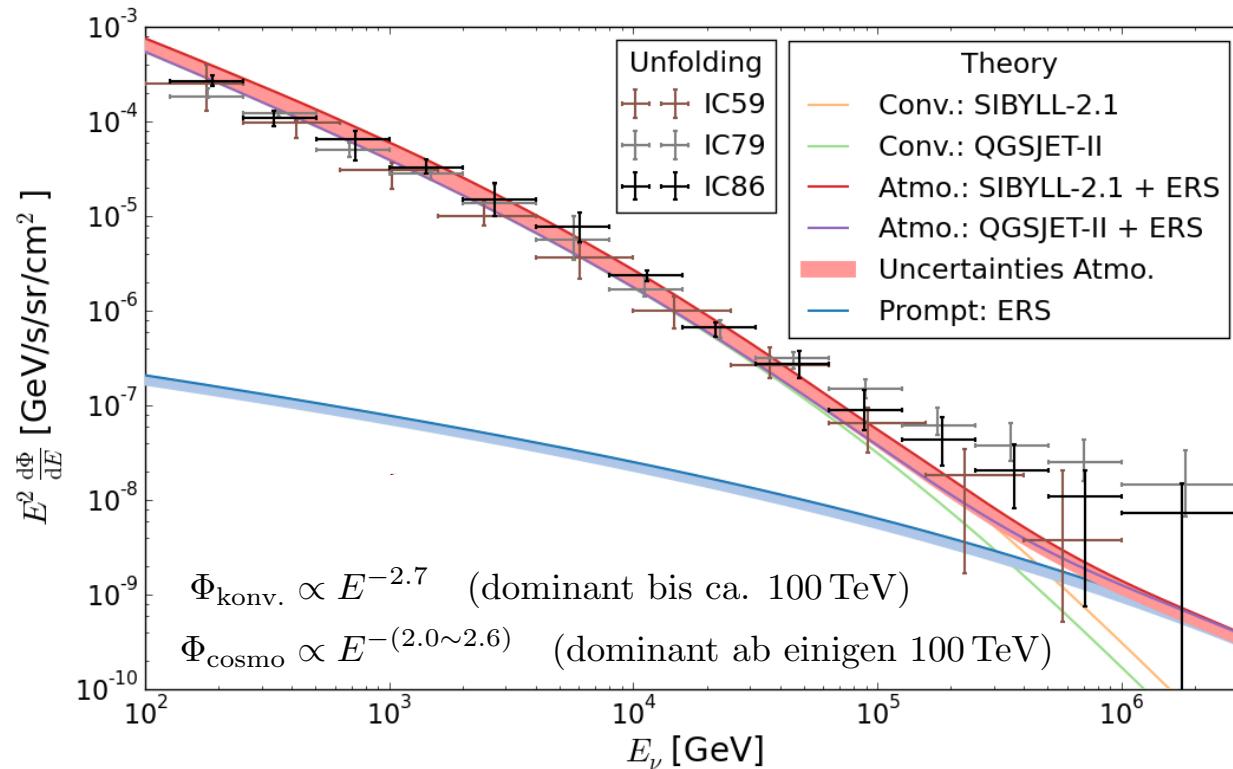
Myon-Neutrinos in IceCube

- Für Neutrino Energien > 100GeV muss zwischen zwei verschiedenen Ursprüngen der Neutrinos unterschieden werden
 - Atmosphärischen Ursprunges: konventionell und prompt
 - Kosmologischen Ursprunges: Genaue Quellen bisher unklar



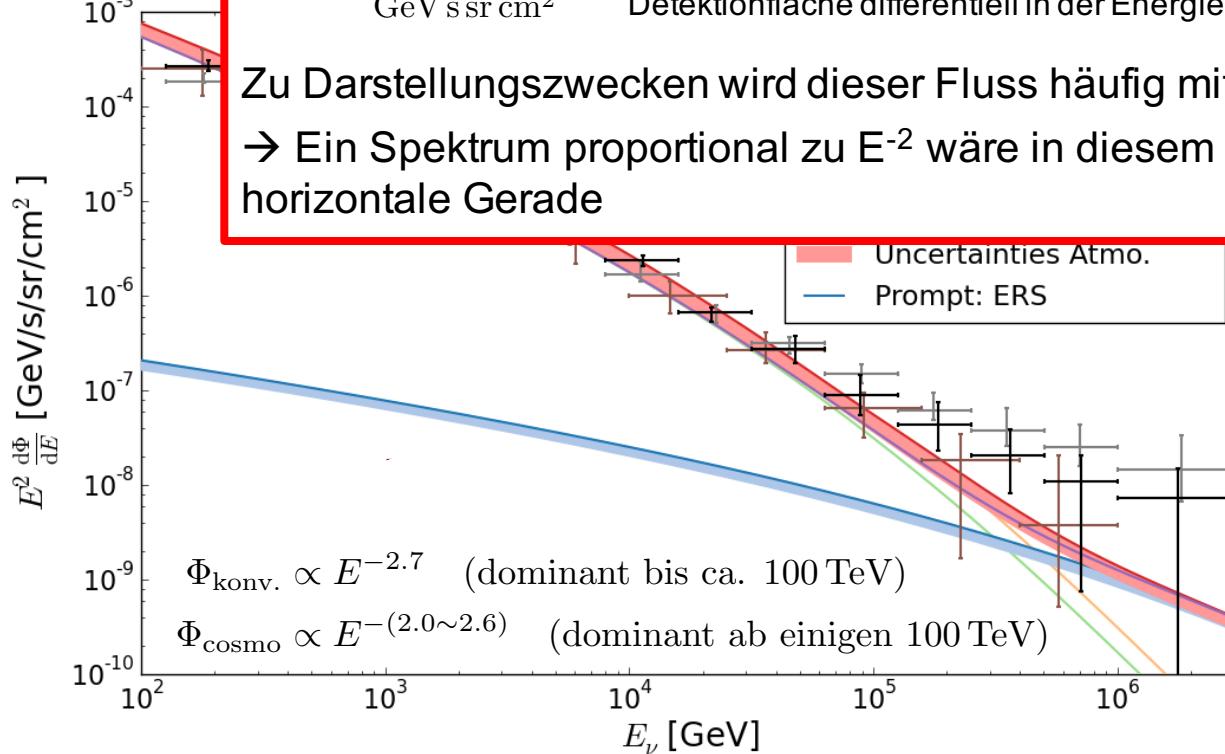
Messung des Energiespektrums

- Gesucht wird das Energiespektrum aller Myon-Neutrinos in IceCube ab einer Neutrinoenergie von > 100 GeV



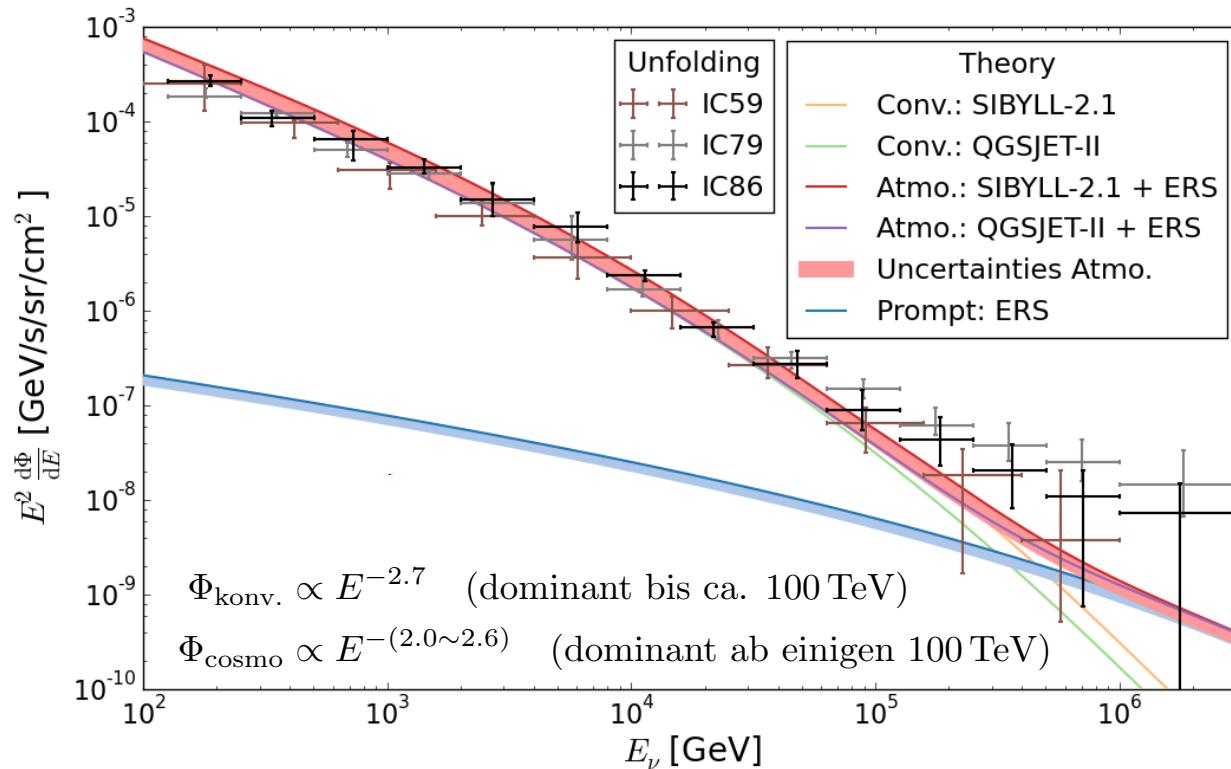
Messung des Energiespektrums

- Gesucht wird das Energiespektrum einer Neutrinoenergie



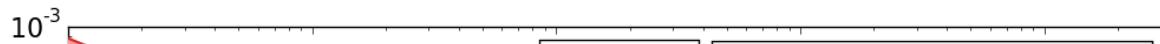
Messung des Energiespektrums

- Gesucht wird das Energiespektrum aller Myon-Neutrinos ins IceCube ab einer Neutrinoenergie von > 100 GeV



Messung des Energiespektrums

- Gesucht wird das Energiespektrum aller Myon-Neutrinos ins IceCube ab einer Neutrinoenergie von $> 100 \text{ GeV}$

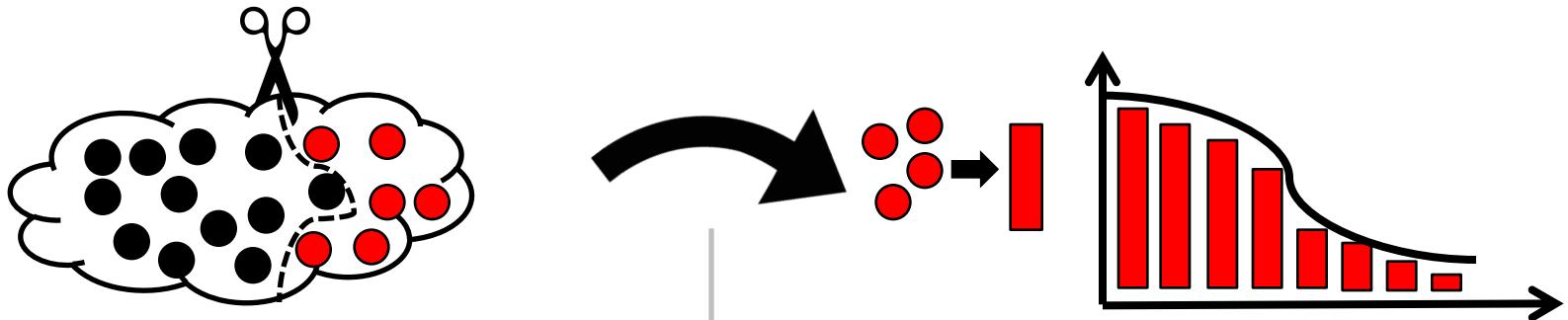


Herausforderungen:

- Der indirekte Nachweis der Myon-Neutrinos durch Myonen (CC-Wechselwirkung) im Detektor
 - Dem Myon im Detektor ist der Ursprung nicht anzusehen. Somit ist eine Unterscheidung zwischen neutrino-induzierten Myonen und atmosphärischen Myonen nicht trivial
- Es kann nur die Energie des Myons im Detektor gemessen werden
 - Physik und Detektoreffekte verschmieren die Energie und machen eine direkte Umrechnung von Myon- zu Neutrinoenergie nicht möglich

Messung des Energiespektrums

- Die Analyse wird in zwei Schritte unterteilt, in denen jeweils eine der beiden Herausforderungen angegangen wird



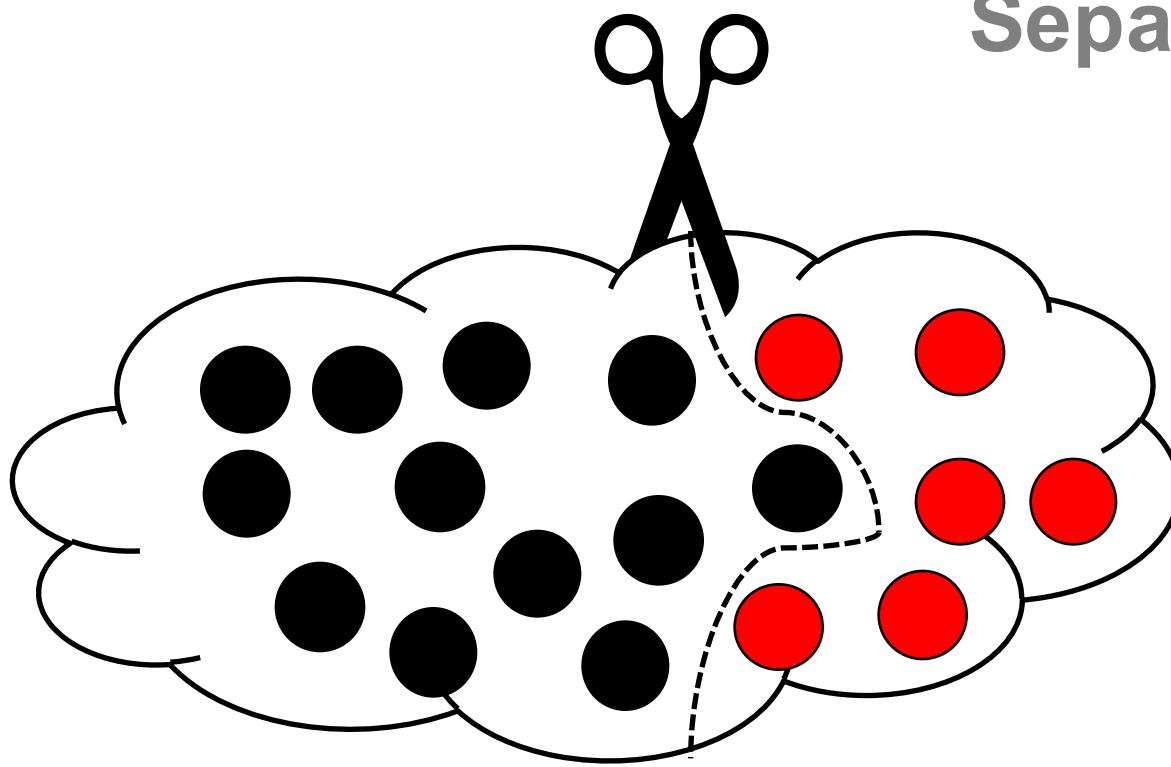
Schritt 1:

- Trennung von Signal-Untergrund
- Im Zentrum stehen hier Methoden des **Data-Minings**
(Thema der SMD bis Weihnachten)

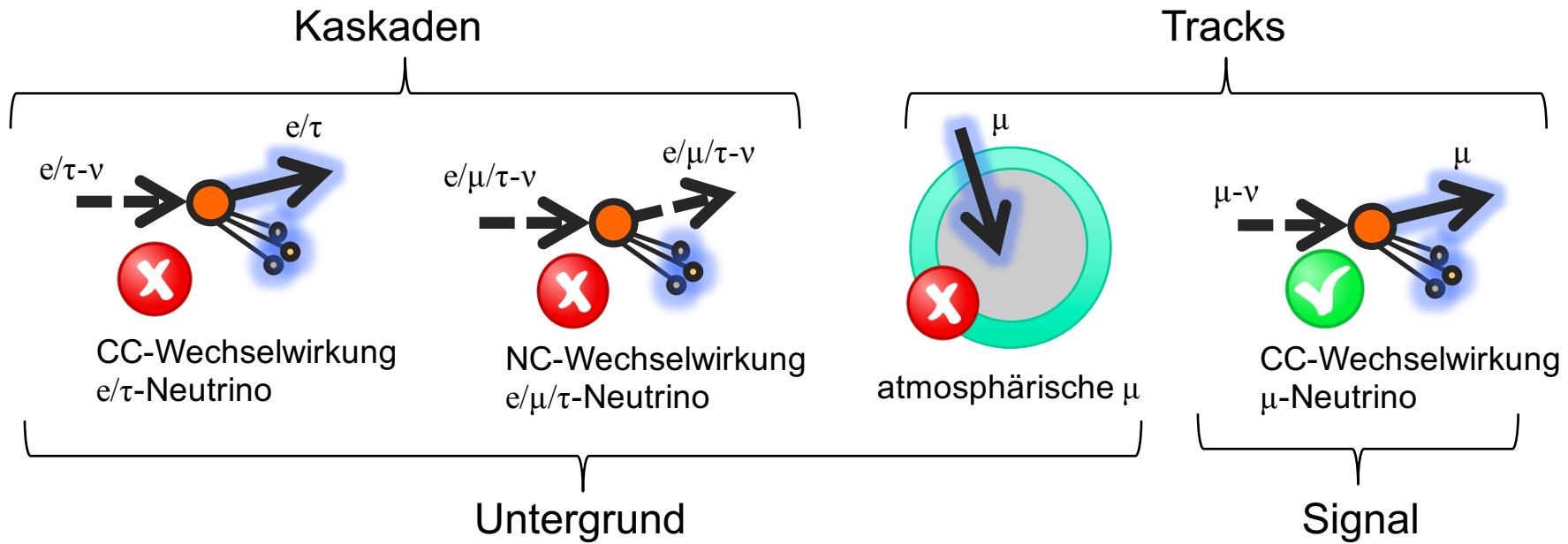
Schritt 2:

- Bestimmung des Spektrums auf getrennten Ereignissen
- Im Zentrum steht hier die **Entfaltung**
(wichtiges Thema der SMD nach Weihnachten)

Schritt 1: Separation



Trennung von Signal und Untergrund



- Alle Kaskaden gehören zum Untergrund
- Bei den Tracks muss zwischen atmosphärischen Myonen und neutrino-induzierten Myonen unterschieden werden

Trennung von Signal und Untergrund – Die Daten

- Die Daten liegen in Form von (großen) Tabellen vor und bringen eine Herausforderung mit sich

hochdimensional
 $M \sim 2000$

$$\frac{\text{Sig}}{\text{Bkg}} \sim \frac{1}{1000000}$$

Simulation (noch)
nicht perfekt

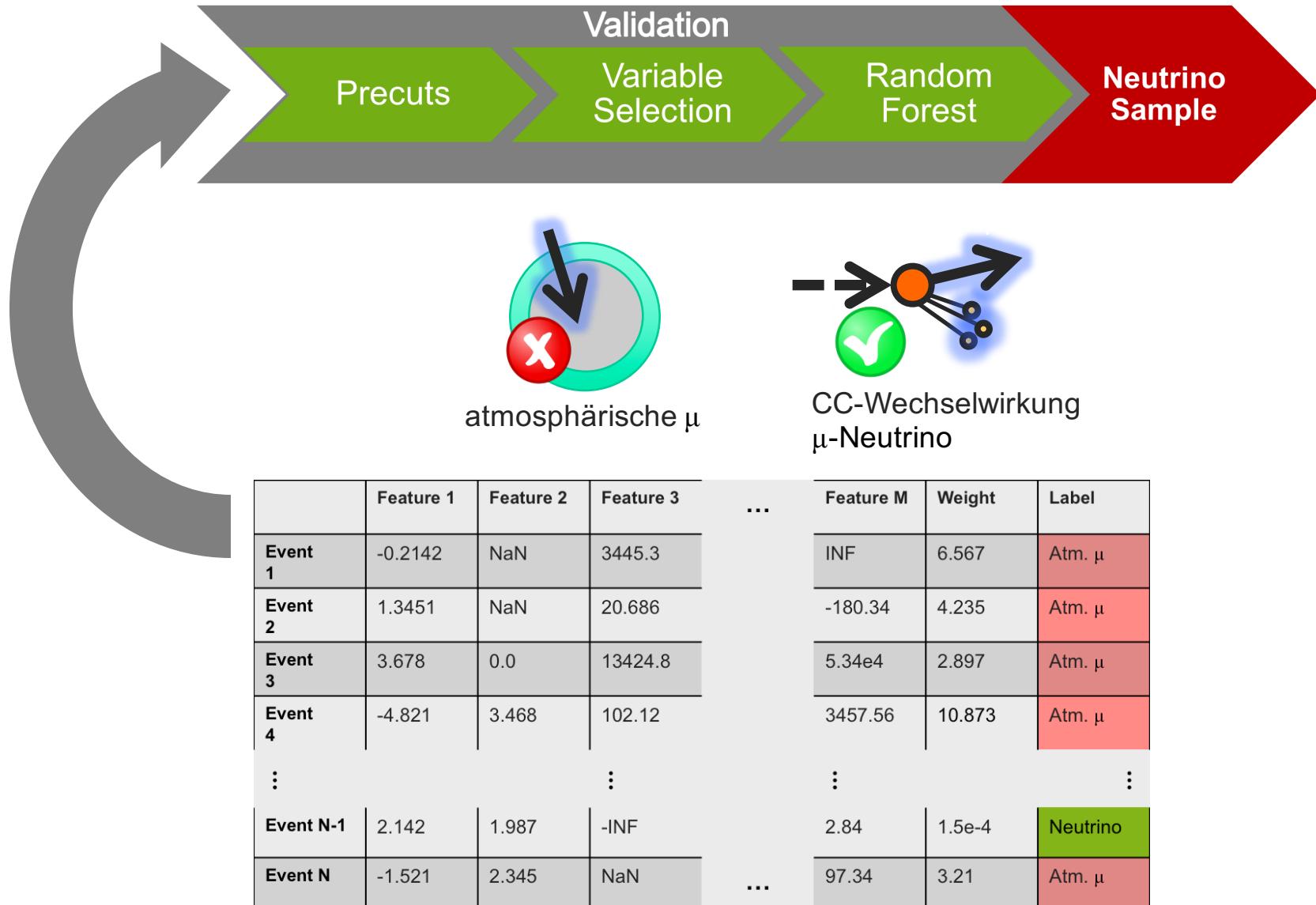
Steiles Spektrum
 $\Phi \propto E^{-2.7}$

	Feature 1	Feature 2	Feature 3
Event 1	-0.2142	NaN	3445.3
Event 2	1.3451	NaN	20.686
Event 3	3.678	0.0	13424.8
Event 4	-4.821	3.468	102.12
:		:	
Event N-1	2.142	1.987	-INF
Event N	-1.521	2.345	NaN

...

	Feature M	Weight	Label
INF	6.567	Atm. μ	
-180.34	4.235	Atm. μ	
5.34e4	2.897	Atm. μ	
3457.56	10.873	Atm. μ	
:			:
2.84	1.5e-4	Neutrino	
97.34	3.21	Atm. μ	

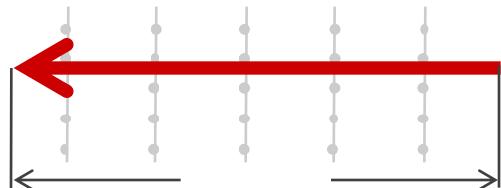
...



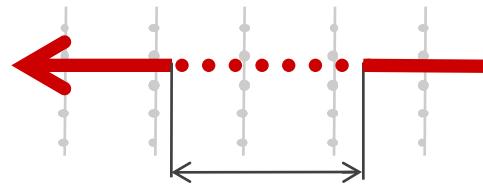


Trennung von Signal und Untergrund

- Schnitte, um Kaskaden und Events schlechter Qualität zu entfernen



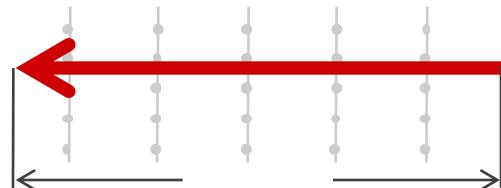
Länge des Tracks > 200 m



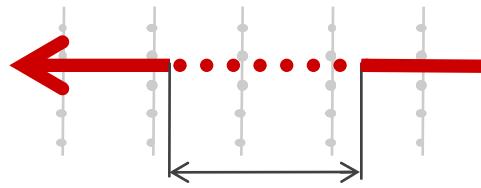
Größe „Lücke“ im
Track < 400 m

Trennung von Signal und Untergrund

- Schnitte, um Kaskaden und Events schlechter Qualität zu entfernen

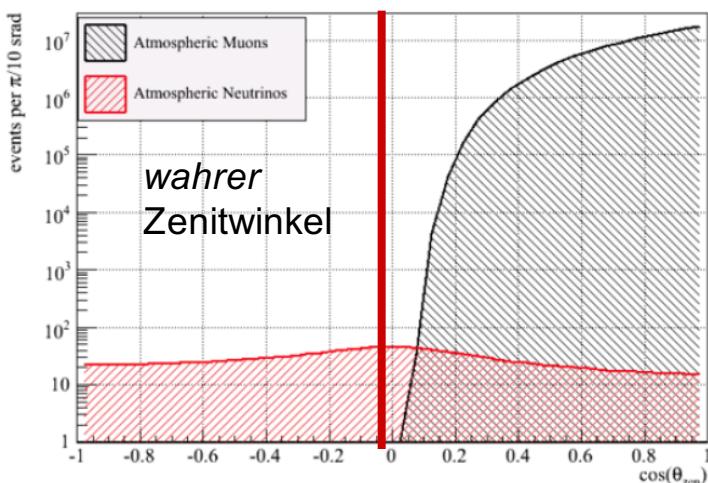


Länge des Tracks > 200 m

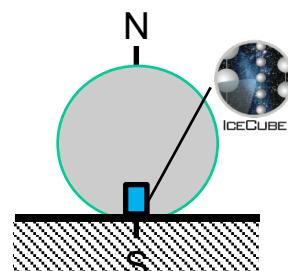


Größte „Lücke“ im
Track < 400 m

- Die Erde als Abschirmung gegen atmosphärische Myonen
 - Eindringtiefe von Myonen ins Eis: ca. 1km pro TeV



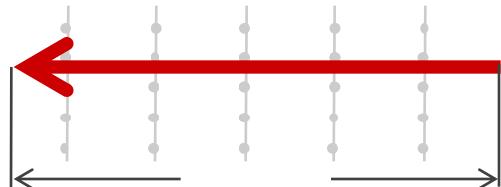
wahrer
Zenitwinkel



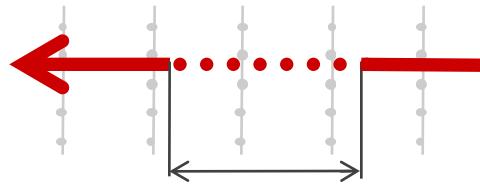
Zenitwinkel > 86 deg

Trennung von Signal und Untergrund

- Schnitte, um Kaskaden und Events schlechter Qualität zu entfernen

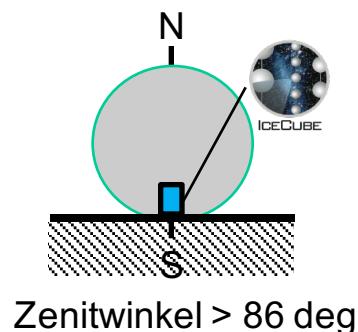
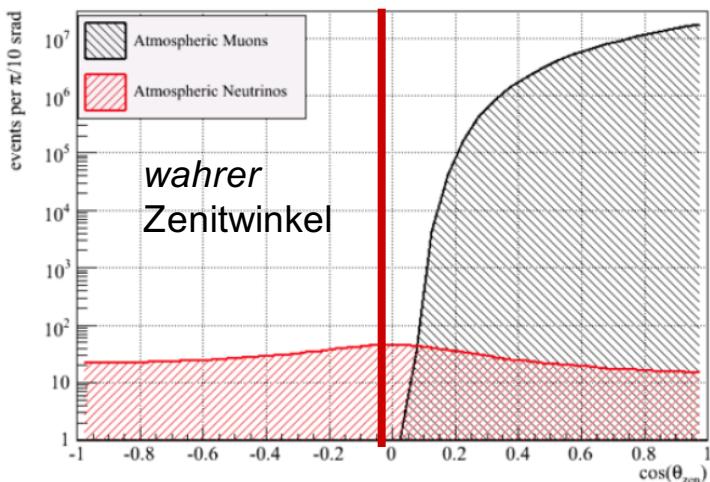


Länge des Tracks > 200 m

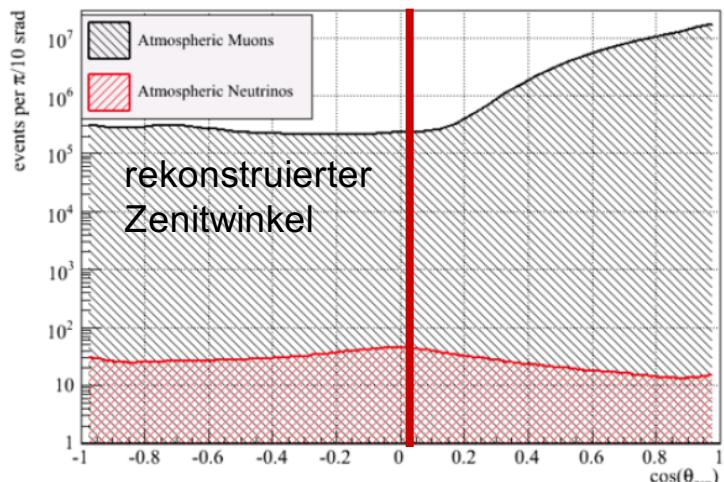


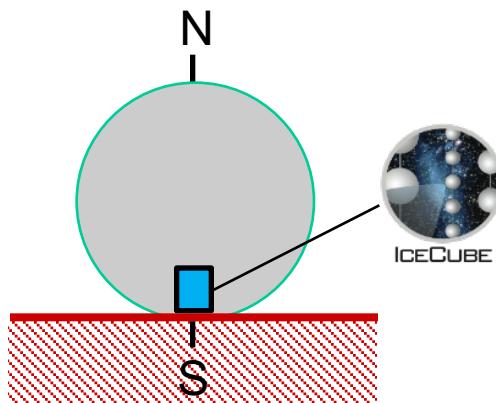
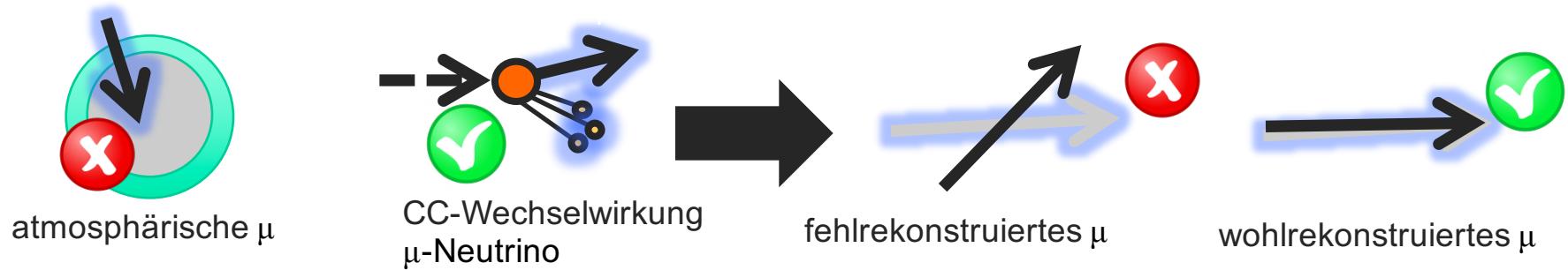
Größte „Lücke“ im
Track < 400 m

- Die Erde als Abschirmung gegen atmosphärische Myonen
 - Eindringtiefe von Myonen ins Eis: ca. 1km pro TeV



Zenitwinkel > 86 deg

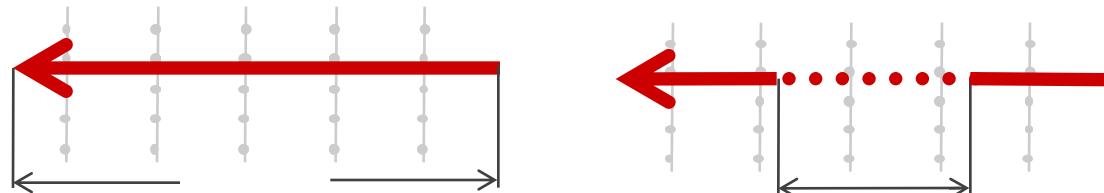


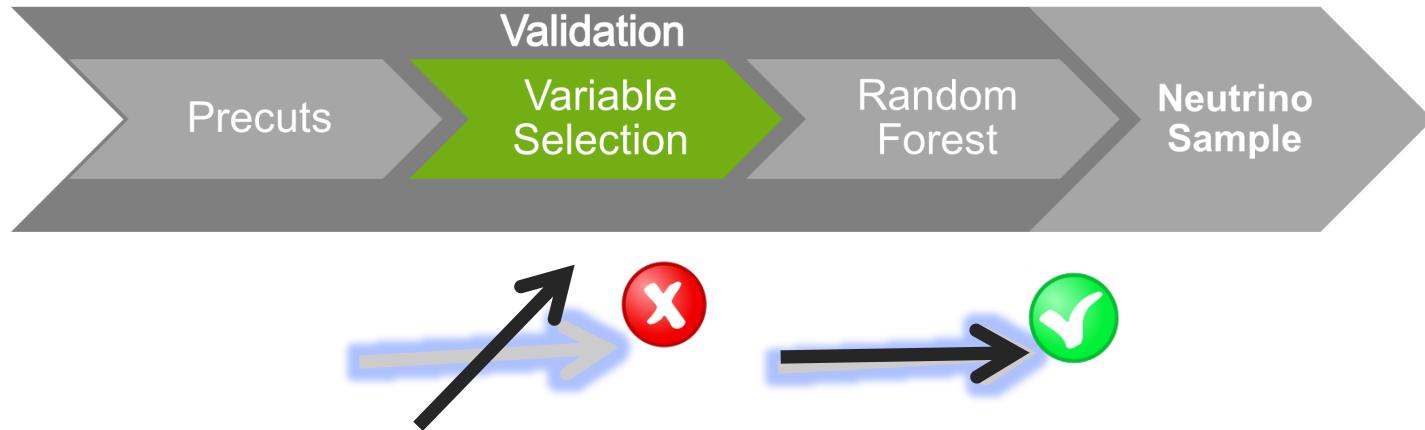


Zenitwinkel > 86 deg

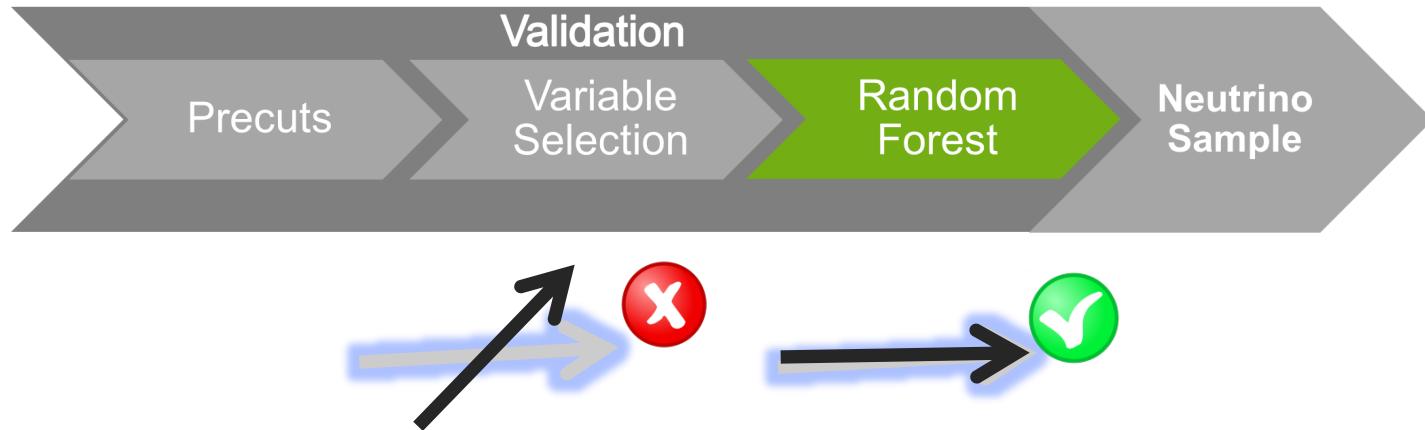
Länge des Tracks > 200 m

„Lücken“ im Track < 400 m

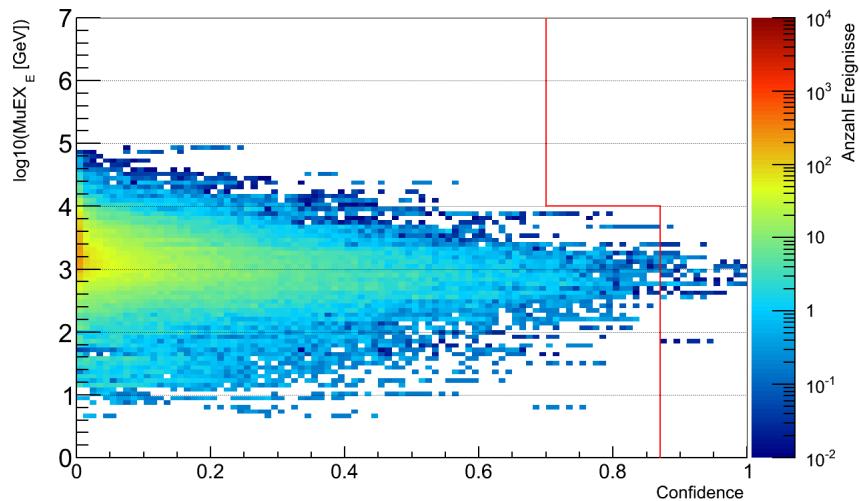
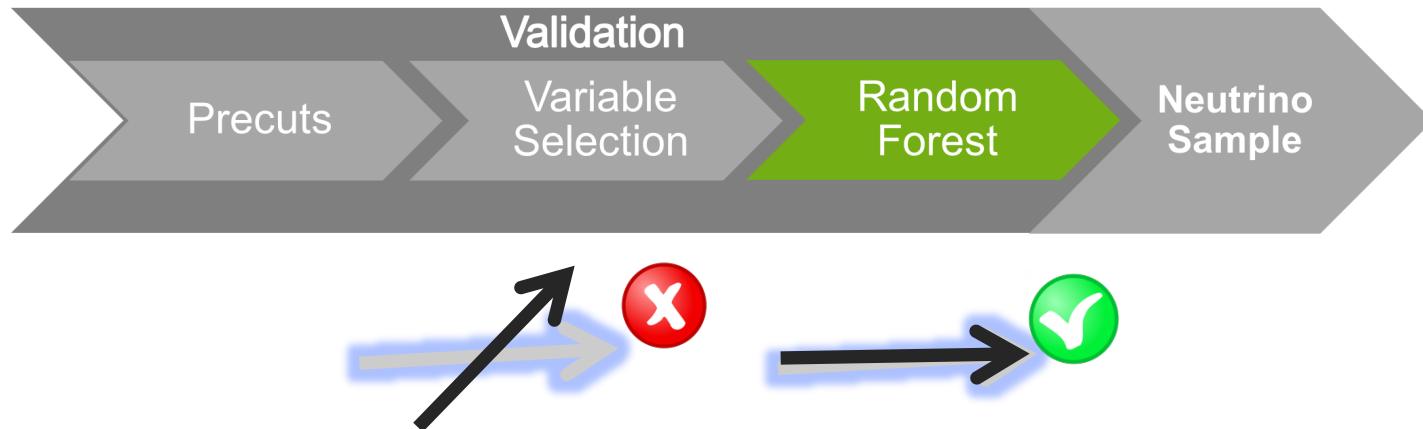




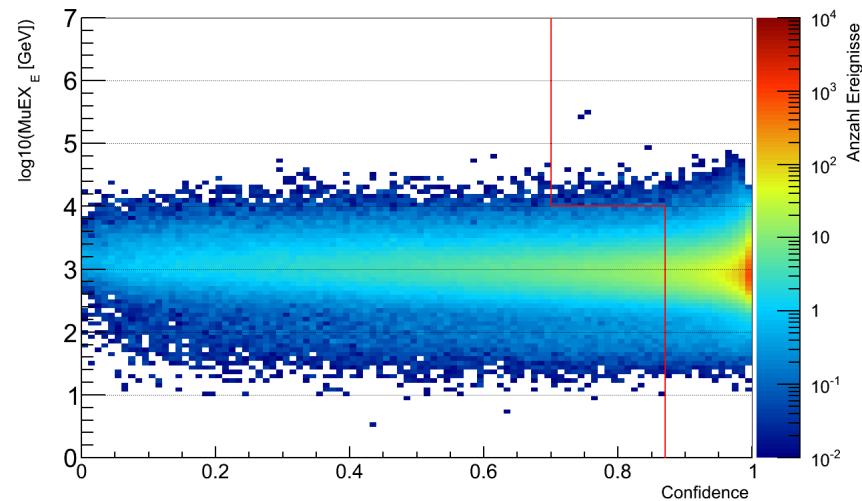
- Hochdimensionaler Datensatz mit vielen redundanten oder sogar unnützen Informationen für die Fragestellung
- Teilweise passen die Simulationen und die Messung in Variablen nicht zu einander
- Viele Dimensionen machen den Datensatz unhandlich und verhindern eine effektive Trennung
 - Auswahl von wenigen Variablen, die das Event möglichst gut beschreiben und viele Informationen über die Art des Events enthalten



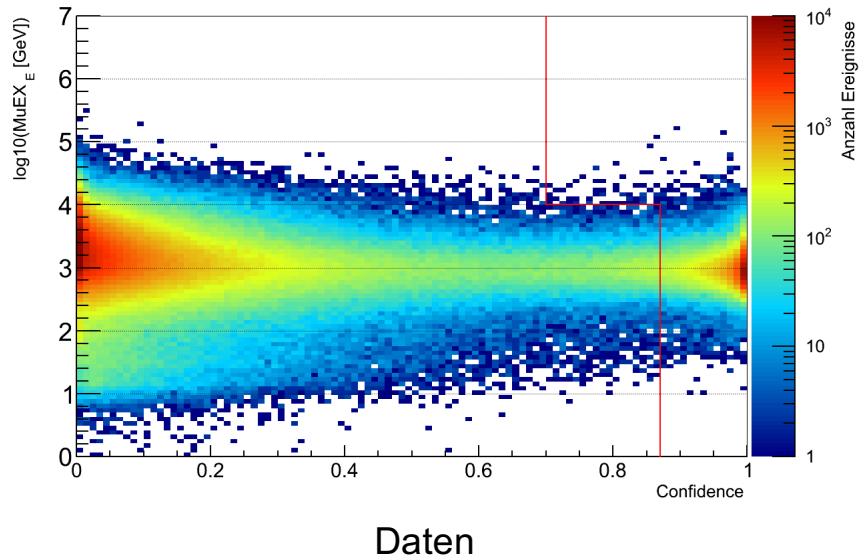
- Gesucht wird ein Modell, dass aus den Strukturen und Zusammenhängen im Datensatz Vorhersagen ableitet, ob es sich beim Event um ein Signal- oder Untergrund-Event handelt
 - Algorithmen des maschinellen Lernens (sogenannt Lernalgorithmen) können auf *gelabelten* Events (z.B. aus Simulationen) trainiert werden und treffen Vorhersagen für Events unbekannter Zugehörigkeit
 - Hier wird ein Random Forest genutzt, dessen Vorhersage (*Confidence*) zwischen 0 und 1 liegt und die Sicherheit angibt mit der der Random Forest das Event als Signal klassifiziert.



atmosphärische Myonen (Simulation)



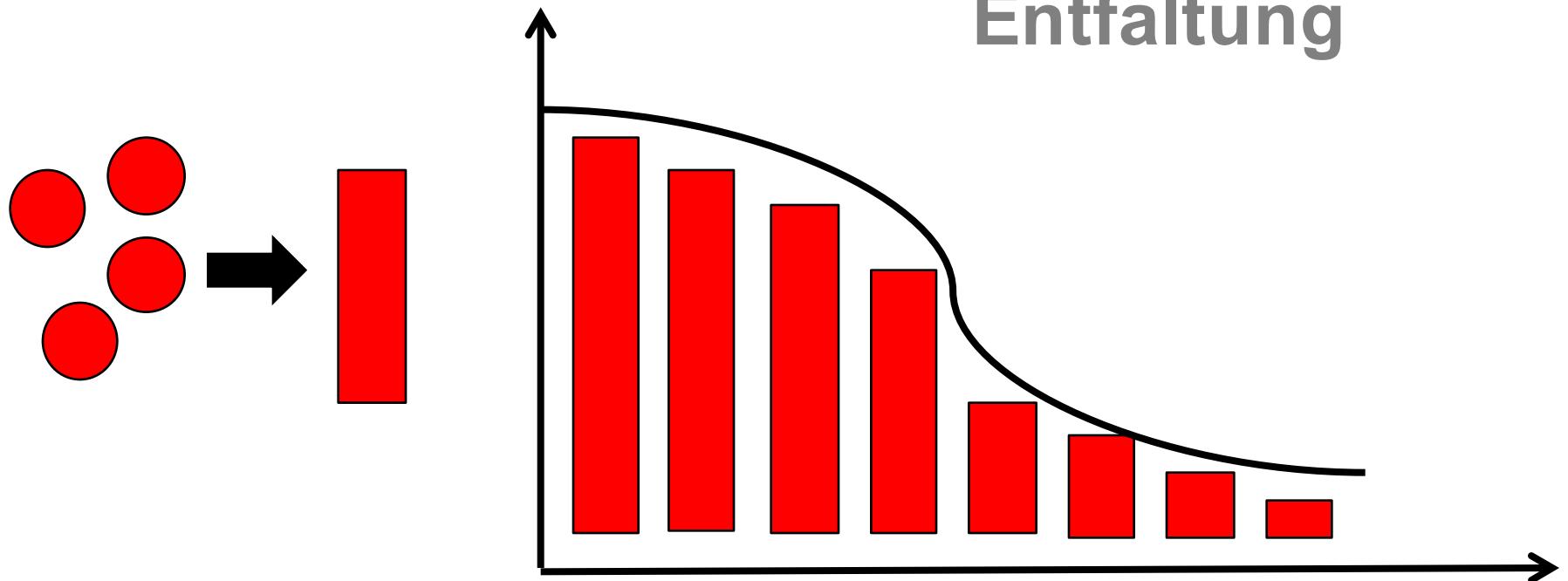
neutrino-induzierte Myonen (Simulation)



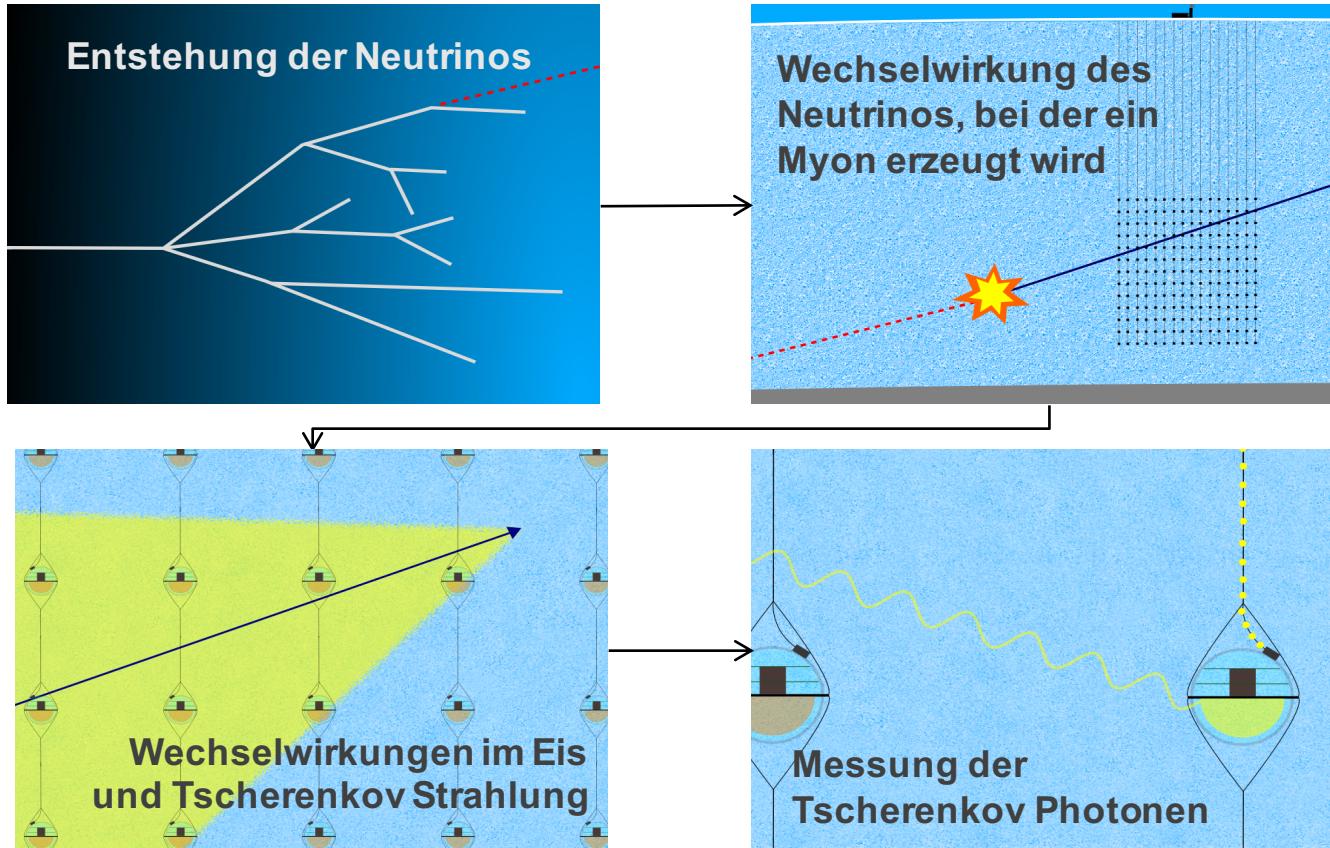
Energie	Atmos. μ	Neutrino μ
Gesamter Bereich	491 ± 263	93280
$E_\nu > 1 \text{ TeV}$	169 ± 140	39088
$E_\nu > 10 \text{ TeV}$	0.01 ± 0.01	233

- $99.6 \pm 0.2\%$ sind tatsächlich Myon-Neutrino Events
- 99.999% des Untergrunds wurde abgelehnt

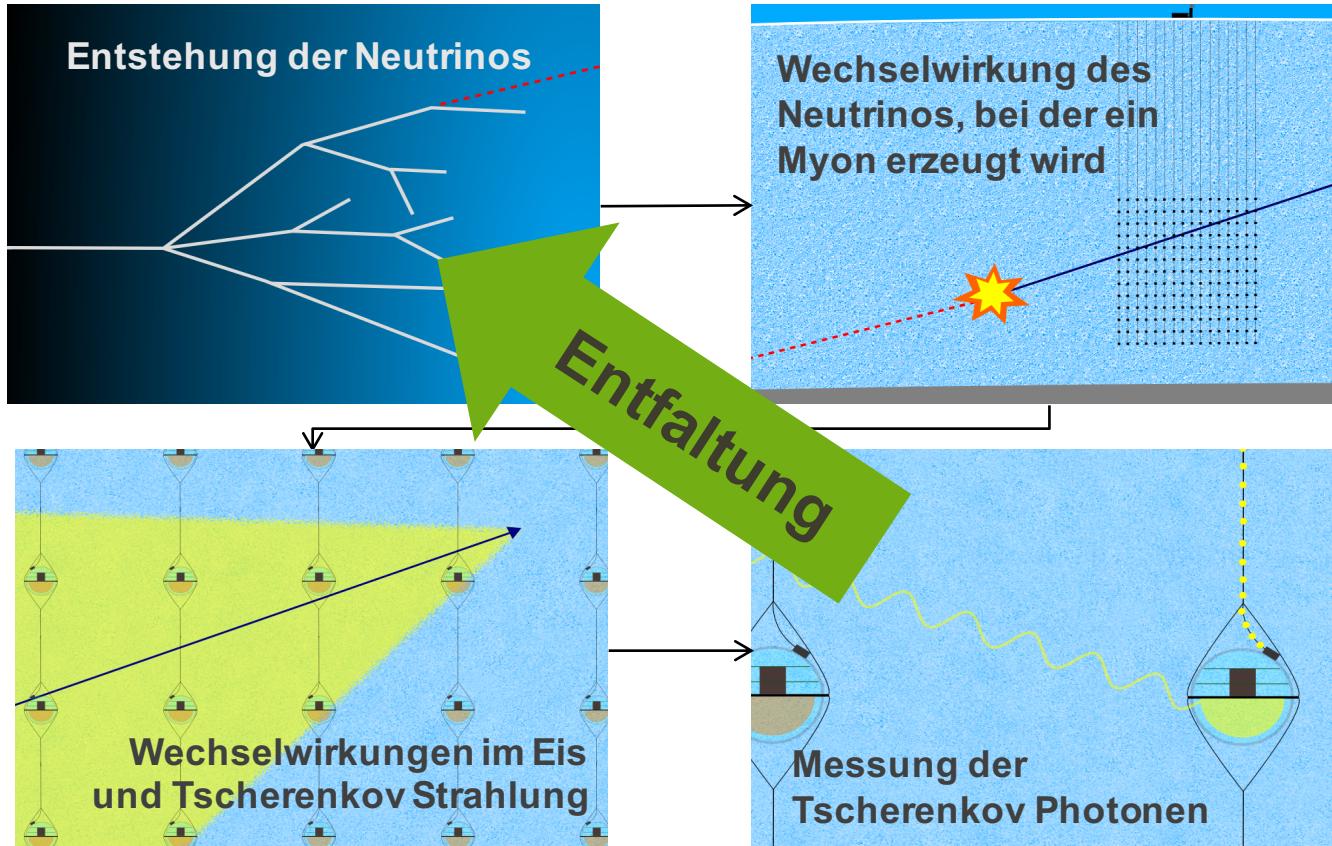
Schritt 2: Entfaltung



Entfaltung

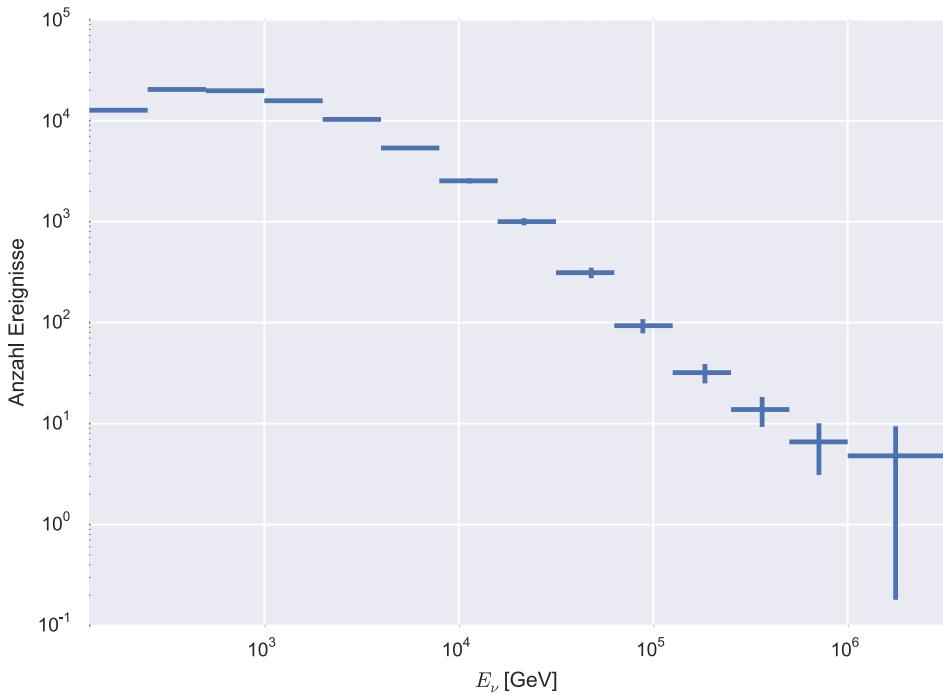


Entfaltung



$$g(y) = \int A(x, y) f(x) dx$$

Entfaltung des Myon-Neutrino Energiespektrums



- Das Ergebnis der Entfaltung ist das Spektrum in Ereigniszahlen
- Mit Hilfe der Simulationen wird eine Akzeptanz-korrektur bestimmt
- Division mit dem beobachteten Himmelsbereich und der Messzeit verwandelt die Ereigniszahlen in einen Fluss
- Durch Vielfaches Entfalten von Simulationen, die systematische Variationen enthalten, werden die systematische Unsicherheiten abgeschätzt

Entfaltung des Myon-Neutrino Energiespektrums

