technische universität
dortmund

IceCube
SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Exploring Optical Properties of Antarctic Ice with IceCube Using Gradient Descent

Alexander Harnisch

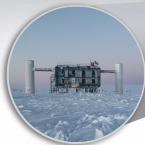**Aachen HAP Workshop 2019**

January 19, 2019

IceCube
SOUTH POLE NEUTRINO OBSERVATORY

50 m

IceTop

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

1450 m

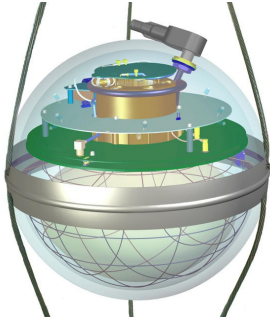86 strings of DOMs, set 125 meters apart
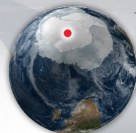
Amundsen–Scott South Pole Station, Antarctica
A National Science Foundation-managed research facility

60 DOMs on each string

DOMs are 17 meters apart

IceCube detector

DeepCore

Digital Optical Module (DOM)
5,160 DOMs deployed in the ice

2450 m

Antarctic bedrock

technische universität dortmund

IceCube
SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information
by Resource-Constrained Data Analysis
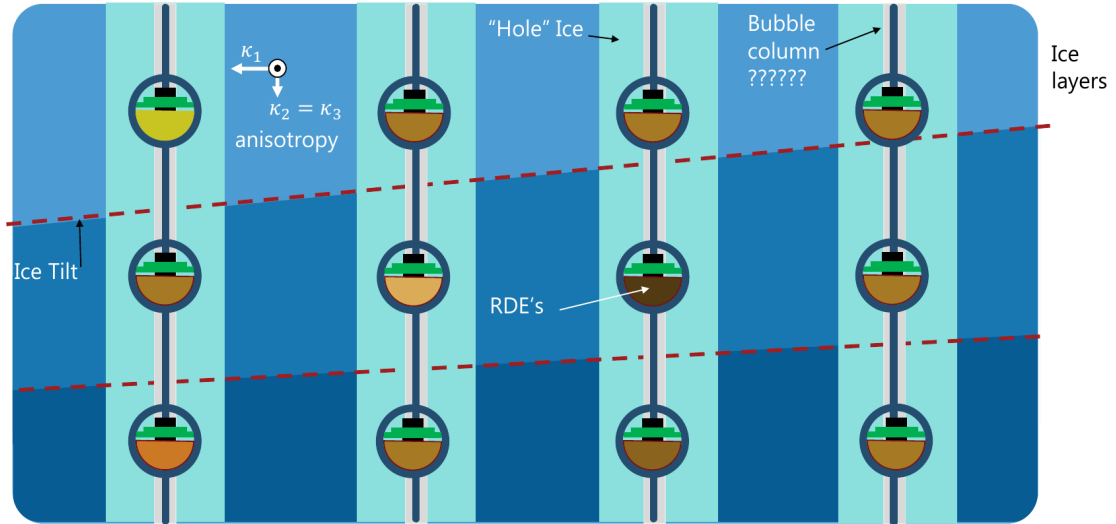
Lehrstuhl für
Experimentelle Physik 5

Introduction

- IceCube detects neutrinos by measuring the Cherenkov light emitted by secondary particles
- For that we need precise simulations of photon propagation within the ice
- For modeling the ice we divide it into layers in z direction
- There are global ice model parameters and parameters for each layer, mainly scattering and absorption coefficients
- There is a flasher board on each DOM which we use as light sources for callibration

The SPICE 3.2 Ice Model

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Photon Propagation in Ice

Two things can happen to a photon while moving through the ice: It can get **scattered** or **absorbed**.

1. Scattering
Distance between successive scattering processes:

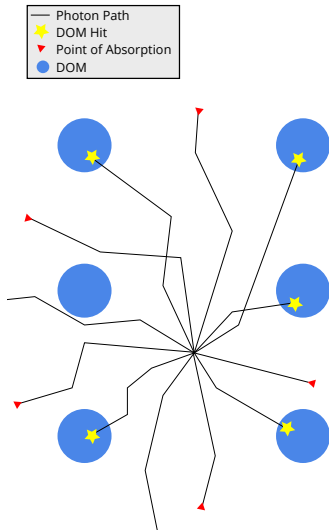$$p(d_{\text{scat}}) = be^{-bd_{\text{scat}}} .\tag{1}$$

We define the effective scattering coefficient $b_e = b(1 - \langle cos\theta \rangle)$ where θ is the deflection angle, which follows a pdf $p(\cos\theta)$

2. Absorption
The travel distance before absorption:
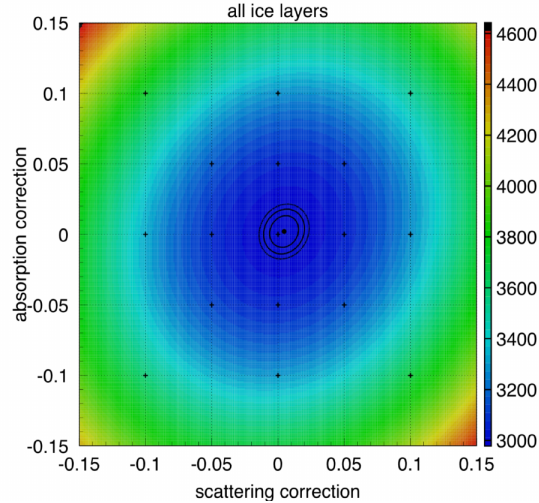
$$p(d_{\text{abs}}) = ae^{-ad_{\text{abs}}} .\tag{2}$$

Photon Propagation Simulation

Legend:
- Photon Path
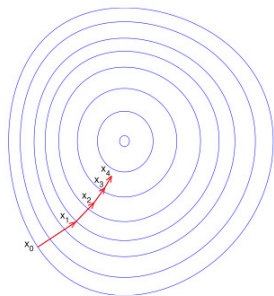- DOM Hit
- Point of Absorption
- DOM

1. Initialize photons
2. While not hit or absorbed
   2.1 Propagate photon
      - Hit? Stop propagation at DOM contact point
      - Absorbed? Stop propagation at point of absorption
      - Otherwise propagate to next scattering point
   2.2 Scatter photon, if not hit or absorbed
3. Output: DOM responses

See pseudo code on backup-slide 28

technische universität
dortmund

IceCube
SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Likelihood Grid Scan

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis
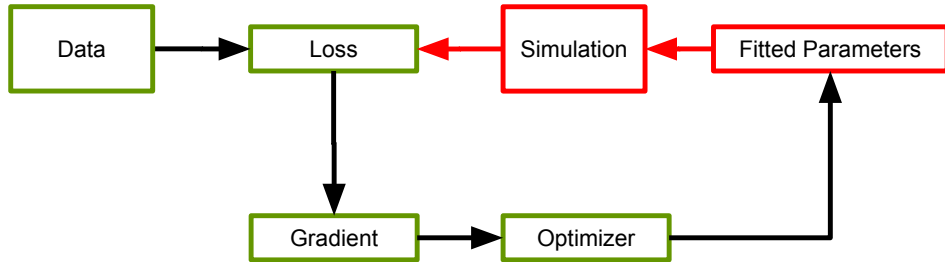
Lehrstuhl für
Experimentelle Physik 5

Motivation

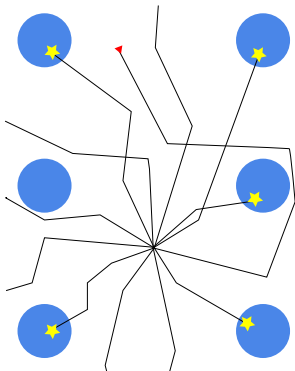- Fitting ice parameters by performing iterative grid searches is extremely complex and time-consuming
- We want to compute gradients of the likelihood, to be able to perform gradient descent instead of grid searches
- This might reduce the cost and improve scalability
- In this talk we focus on bulk ice absorption coefficients
- We should at least be able to verify the current best fit using this independent method

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

General Idea

Challenge: Making the red part differentiable.

technische universität dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Photon Propagation Simulation - Modified to Allow Differentiation

— Photon Path
★ DOM Hit
▼ Cut-off point
● DOM
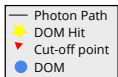
1. Initialize photons
2. While not hit or cut-off reached
   2.1 Propagate photon
     - Hit? Stop propagation at DOM contact point
     - Cut-off reached? Stop propagation
     - Otherwise propagate to next scattering point
     - **Log traveled distance in each layer**
   2.2 Scatter photon, if not hit or cut-off
3. Intermediate output for each hit: The DOM that was hit and the traveled distance in each layer

Cut-off based on total travel distance and ignore all photons, that were cut-off.

technische universität
dortmund

ICECUBE
SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Calculate DOM Response from Intermediate Result

Input: Photon travel distances D_j for each layer j and which DOM was hit for each Photon, that hit a DOM.

Calculate the probability for each photon to reach the DOM, after propagation:

$$p_{\text{Hit}} = 1 - p_{\text{Absorbed}} = \exp\left(-\int_0^D a(x)\mathrm{d}x\right)$$

$$\overset{\text{Layer}}{=} \exp\left(-\sum_{j=1}^{N_{\text{Layer}}} a_j D_j\right) \tag{3}$$

where a_j is the absorption coefficient of the j-th layer.

Output: Summed hit probabilities of all photons for each DOM

$$s_i = \sum_{j=1}^{N_{\text{Hits},i}} p_{\text{Hit},ij} \tag{4}$$

technische universität
dortmund

IceCube
SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Getting the Gradient

The loss F we are trying to minimize is the negative logarithm of the likelihood ratio
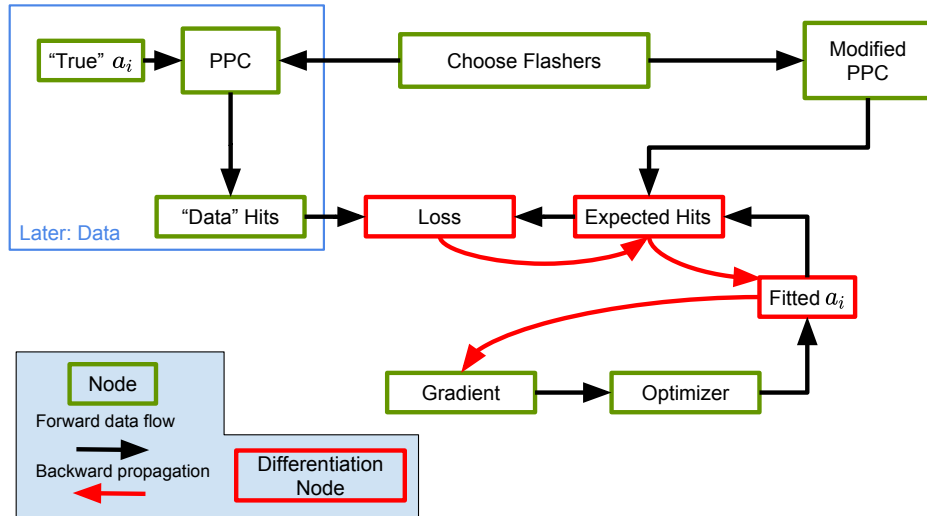
$$-F = \log \mathcal{L}_{\text{Ratio}} \,. \tag{5}$$

We now have a direct chain between a suitable loss function F and absorption coefficients whitout any control statements, which we can use to compute the gradient

$$\nabla_{\vec{a}} F = \cdots \tag{6}$$

by using TENSORFLOW's automatic differentiation (backpropagation). We could also define the gradient by hand, which would be simple but tedious.
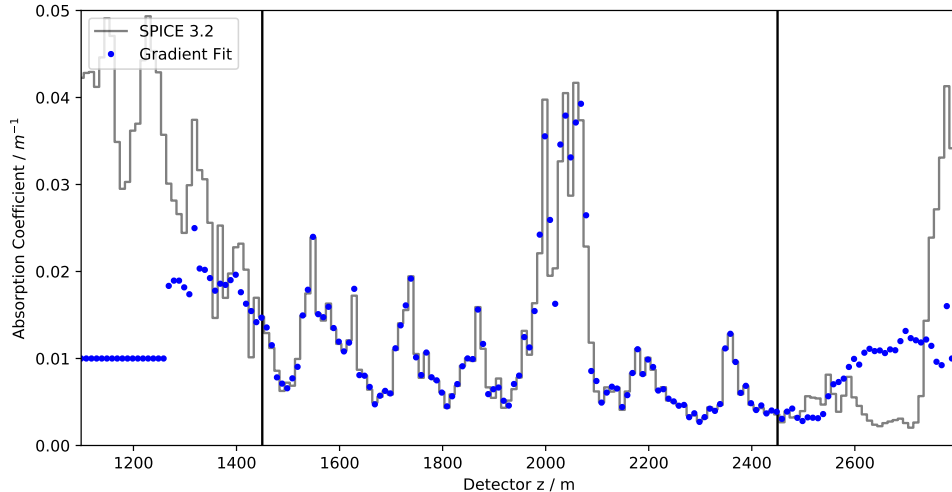
technische universität dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Fitting Algorithm

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Video of the Learning Progress

Hyperlink to video

technische universität dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Result for fake data

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Loss

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Conclusion and Outlook

Seems to be working! We should at least be able to verify the current fit with this independent approach.

Problems:

- Anisotropy
- Possibly other differences between data and simulation, that might require regularization
- Performance: Copying stuff around GPU and CPU memory. Need for general optimization

Ideas:

- Scattering: Many possible approaches
  - Assuming a strong correlation between absorption and scattering
  - Hybrid between grid search and gradient descent approach
  - Estimating the gradient for scattering coefficients, e.g. by resampling the arrival time for each photon
  - **Use a DNN!** More in Sebastian's talk
- Anisotropy: Tessellated sphere idea proposed by Martin Rongen
- Possibly do hardware simulation
- Use timing information

Next? Get it working on real data and incorporate scattering.

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Backup

technische universität dortmund

IceCube

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Optimization Techniques

Playing around with two approaches so far:

1. **DOM Batches**
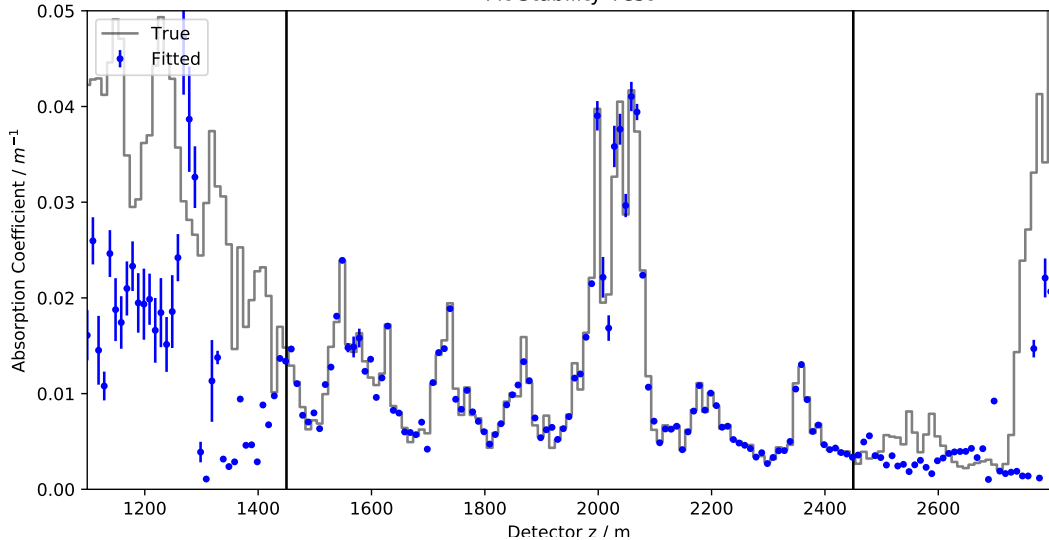   Evaluate the gradient for each flasher event and directly feed it to the optimizer.

2. **Gradient Averaging**
   Evaluate the gradient for each flasher event, but accumulate the gradients for all DOMs on a string and then feed the accumulated gradient to the optimizer. This makes the loss much smoother and steps performed on the same string are directly comparable.
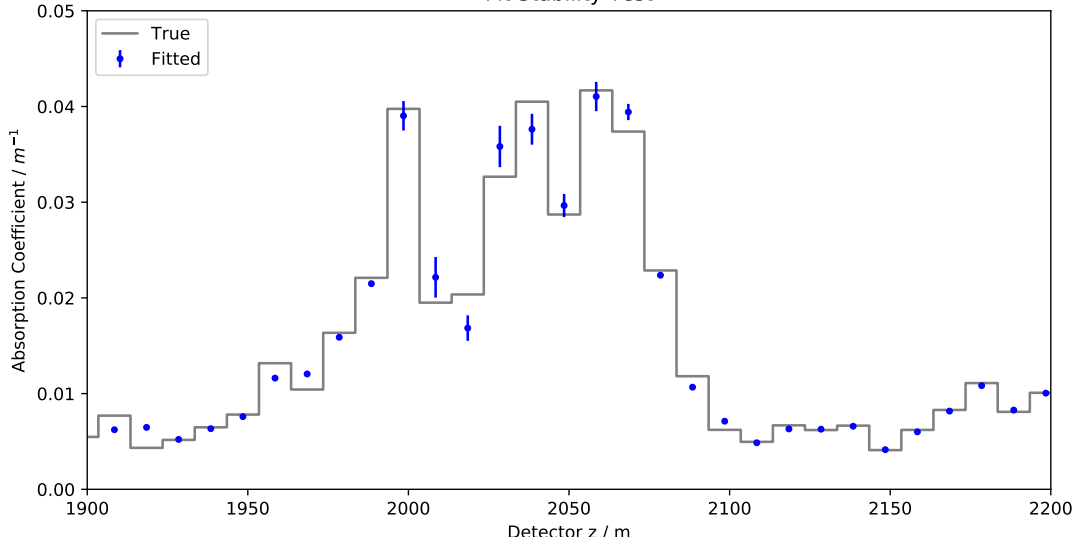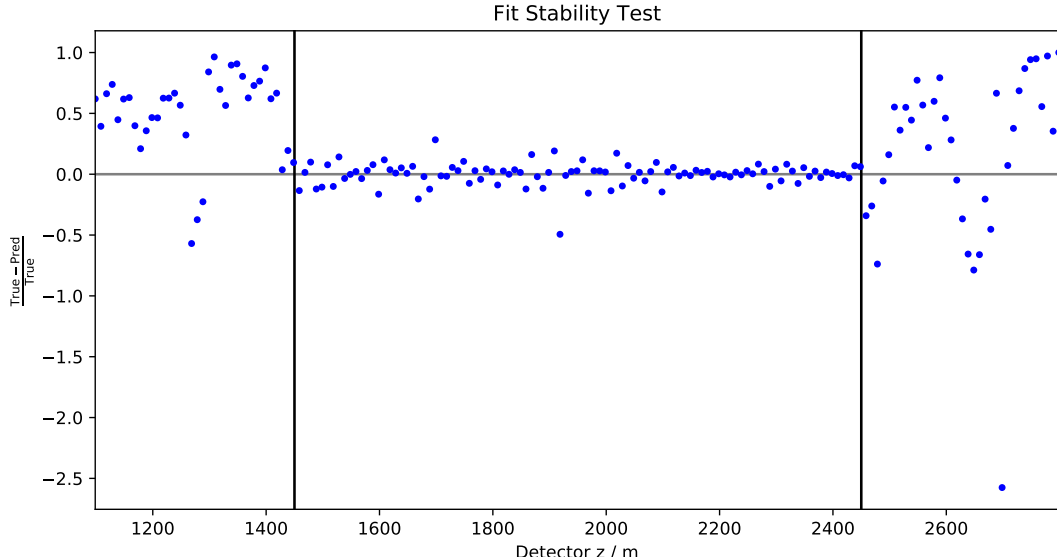
Fit Stability Test

- Idea: Fit multiple times with random initial values to check for stability
- 8 runs with initial coefficients sampled uniformly between $0.005\,\mathrm{m}^{-1}$ and $0.03\,\mathrm{m}^{-1}$
- The following plots show the mean and standard deviation of the mean of those 8 fit results
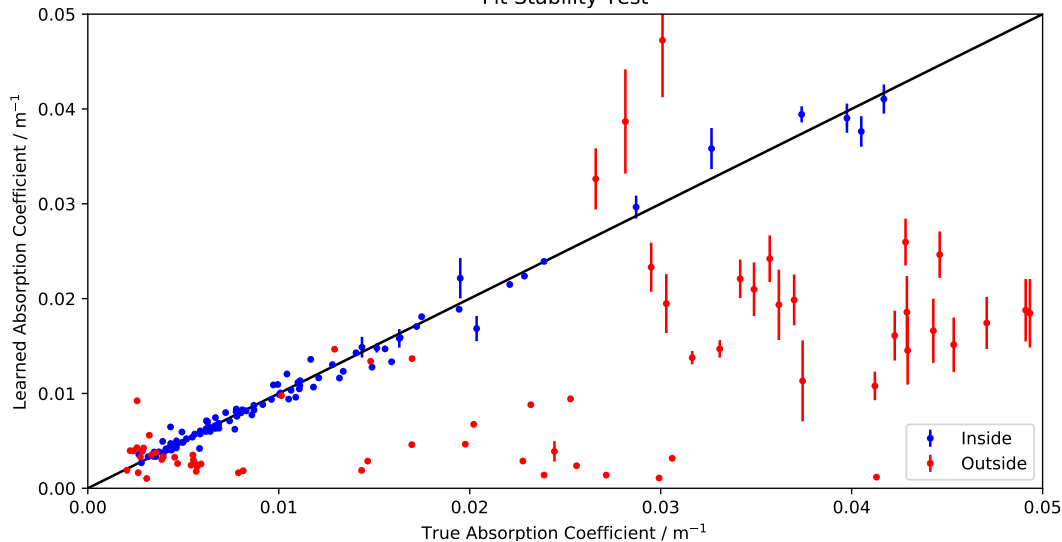
Fit Stability Test

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Fit Stability Test

technische universität
dortmund

IceCube
SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

es

Fit Stability Test

technische universität
dortmund

IceCube
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Fit Stability Test

technische universität dortmund

IceCube
SOUTH POLE NEUTRINO OBSERVATORY

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

The Problem with Differntiating the Propagation Loop

- Automatic differentiation works great for arbitrarily complex programs and can deal with control statements.
- There is one caveat: Those control statements must not depend on the target parameters of the differentiation.
  - In that case the structure of the program changes depending on the target parameter, information about other branches of the program is not included in a mathematical derivative of the one branch that was executed.
  - It can still work in some cases (e.g. our first results) but generally there is no guarantee for the gradient to point in the right direction.

technische universität dortmund

ICECUBE
South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

The Problem - Super Simple Example

Consider the following program:

```python
def y(x):
    if x == 2:
        return 4
    else:
        return 2*x
```

- In case of $x \neq 2$ automatic differentiation will provide the correct derivative of $\frac{\partial y}{\partial x} = 2$. In case of $x = 2$ the derivative is 0, but we need it to be 2 in that case as well.
- Automatic differentiation simply evaluates the chain rule along the executed branch of the program. This is only fine as long as the branching does not depend on target parameters.
- In this case the problem could be easily resolved by defining the gradient manually or rewriting the program.
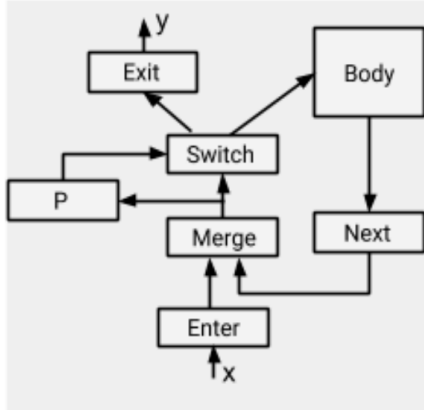
The Problem - Our Case

Our case looks like this (simplified pseudo code):

```python
def simulate(photon, l_abs, l_scat):
    d_abs = sample_absorption(l_abs)
    while d_abs > 0 and not hit:
        d_scat = sample_scattering(l_scat)
        hit = check_for_hits(photon, d_scat)
        if hit:
            propagate_to_hit(photon)
        else:
            if d_abs - d_scat > 0:
                propagate(photon, d_scat)
                scatter(photon)
                d_abs -= d_scat
            else:
                propagate(photon, d_abs)
                d_abs = 0
```
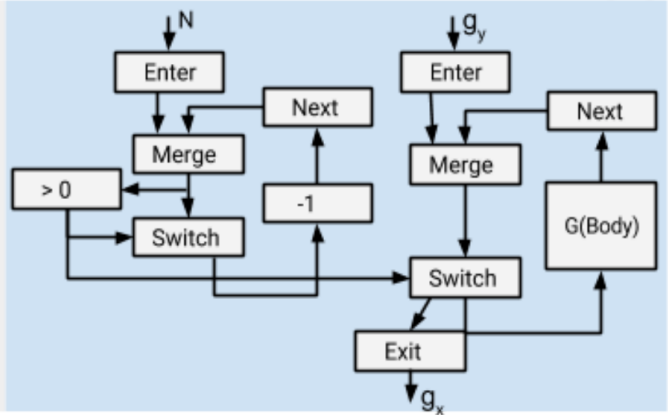
- During forward propagation Tensorflow counts the number of loop iterations. When backpropagating the number of loop iterations is therefore a constant.
- **This means the gradient "does not know" that changing a scattering or absorption length a little bit would lead to one more/less scattering process, even though the derivative of a single scattering process is correct.**
- More on this is explained in the reference given on slide 29 and in a demonstration we wrote.

technische universität dortmund

IceCube South Pole Neutrino Observatory

SFB 876 Providing Information
by Resource-Constrained Data Analysis

Lehrstuhl für
Experimentelle Physik 5

Tensorflow While Loop - Problem

Source: http://download.tensorflow.org/paper/white_paper_tf_control_flow_implementation_2017_11_1.pdf

Problem: N is a constant during backprop. P must not depend on target variables, but it does in our case.