Optical Bistability As Neural Network Nonlinear Activation Function

Davide Bazzanella

20th March 2018

Università degli studi di Trento

All-Optical Neural Networks

2018-03-09

Optical Bistability As Neural Network Nonlinear Activation Function

Davide Bazzanella

20th March 2018

Università degli studi di Trento

Introduction

Introduction

Applying integrated photonics to artificial neural networks architecture design

Develop simulations on standard software libraries that help performance comparisons

All-optical Artificial Neural Networks

Applying integrated photonics to artificial neural networks architecture design

Develop simulations on standard software libraries that help performance comparisons

Outline

2018-03-09

Outline

Introduction
Artificial Neural Networks
Microring Resonator
ANN Simulations
Conclusion

Outline

Outline

Introduction
Artificial Neural Networks
Microring Resonator
ANN Simulations
Conclusion

Outline

Introduction

Artificial Neural Networks

Microring Resonator

ANN Simulations

Conclusion

Outline
Introduction
Artificial Neural Networks
Microring Resonator
ANN Simulations
Conclusion

Outline

Introduction

Artificial Neural Networks

Microring Resonator

ANN Simulations

Conclusion

Outline

Introduction

Artificial Neural Networks

Microring Resonator

ANN Simulations

Conclusion

Outline

Introduction

Artificial Neural Networks

Microring Resonator

ANN Simulations

Conclusion

Outline

Introduction
Artificial Neural Networks
Microring Resonator
ANN Simulations
Conclusion

Artificial Neural Networks

Artificial Neural Networks

ANNs

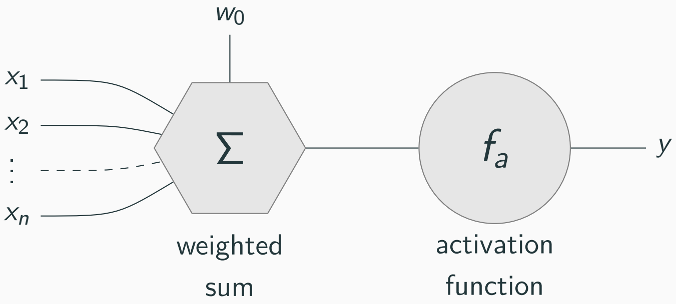
Artificial Neural Networks are computation systems, composed by a collection of nodes that work seemingly biological neurons.

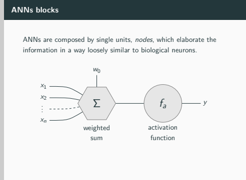
ANNs

Artificial Neural Networks are computation systems, composed by a collection of nodes that work seemingly biological neurons.

ANNs blocks

ANNs are composed by single units, *nodes*, which elaborate the information in a way loosely similar to biological neurons.





What can they do?

What can they do?

What can they do?

What can they do?

ANNs can solve complex problems:

- classification
- clustering
- pattern recognition
- time series prediction

What can they do?

ANNs can solve complex problems:

- classification
- clustering
- pattern recognition
- time series prediction

What can they do?

ANNs can solve complex problems:

- **classification**
- clustering
- pattern recognition
- time series prediction

figures/foo.png

What can they do?
ANNs can solve complex problems:
- classification
- clustering
- pattern recognition
- time series prediction
figures/foo.png

What can they do?

ANNs can solve complex problems:

- classification
- **clustering**
- pattern recognition
- time series prediction

figures/foo.png

What can they do?

ANNs can solve complex problems:

- classification
- clustering
- pattern recognition
- time series prediction

figures/foo.png

What can they do?

ANNs can solve complex problems:

- classification
- clustering
- **pattern recognition**
- time series prediction

figures/foo.png

What can they do?
ANNs can solve complex problems:
- classification
- clustering
- pattern recognition
- time series prediction
figures/foo.png

What can they do?

ANNs can solve complex problems:

- classification
- clustering
- pattern recognition
- time series prediction

figures/foo.png

What can they do?

ANNs can solve complex problems:

- classification
- clustering
- pattern recognition
- time series prediction

figures/foo.png

How do they work?

How do they work?
How do they work?

How do they work?

ANNs can obtain arbitrary decision regions[1]

The amount of free parameters in an ANN, allow .. ?

[1] R. O. Duda et al., *Pattern classification*, (John Wiley & Sons, 2012)

How do they work?

ANNs can obtain arbitrary decision regions[1]

The amount of free parameters in an ANN, allow .. ?

[1] R. O. Duda et al., *Pattern classification*, (John Wiley & Sons, 2012)

How do they work?

- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?
- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?

- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?
- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?

- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?
- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?

- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?
- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?

- training
  - evaluate loss
  - adjust parameters
- validation
- test

How do they work?

- training
  - evaluate loss
  - adjust parameters
- validation
- test

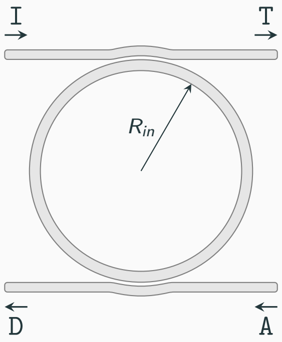
Microring Resonator

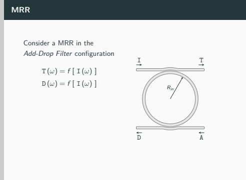
MRR

Consider a MRR in the
Add-Drop Filter configuration

$$\mathtt{T}(\omega) = f\,[\ \mathtt{I}(\omega)\ ]$$

$$\mathtt{D}(\omega) = f\,[\ \mathtt{I}(\omega)\ ]$$



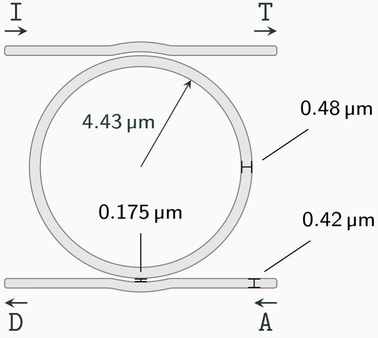


MRR

Consider a MRR in the
Add-Drop Filter configuration

$$\mathtt{T}\,(\omega) = f\,[\ \mathtt{I}\,(\omega)\ ]$$

$$\mathtt{D}\,(\omega) = f\,[\ \mathtt{I}\,(\omega)\ ]$$



MRR
Consider a MRR in the
Add-Drop Filter configuration
$\mathtt{T}(\omega) = f[\mathtt{I}(\omega)]$
$\mathtt{D}(\omega) = f[\mathtt{I}(\omega)]$

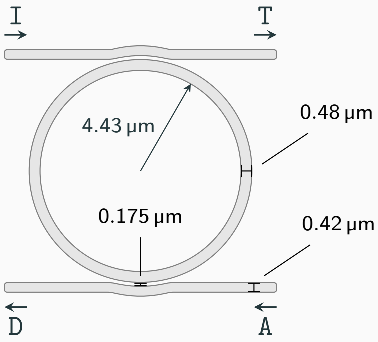
MRR

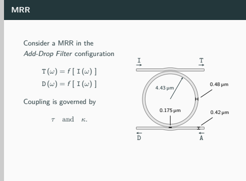
Consider a MRR in the
Add-Drop Filter configuration

$$\mathtt{T}(\omega) = f\,[\,\mathtt{I}(\omega)\,]$$
$$\mathtt{D}(\omega) = f\,[\,\mathtt{I}(\omega)\,]$$

Coupling is governed by

$$\tau \quad \text{and} \quad \kappa.$$





Theory

Linear

2018-03-09

Theory

Linear

Theory

Nonlinear

Theory

Nonlinear

Experiments

2018-03-09

All-Optical Neural Networks
└─Microring Resonator
└─Experiments

Experiments

ANN Simulations

ANN Simulations

Simulation Framework

What means simulating? PyTorch library

Simulation Framework

What means simulating? PyTorch library

Fundamental blocks

model ($FF[f_a]$)

loss criteria (CEL)

weight update criteria (SGD)

Fundamental blocks
model ($FF[f_a]$)
loss criteria (CEL)
weight update criteria (SGD)

model ($FF[f_a]$)

model ($FF[f_a]$)

Cross-Entropy Loss (also known as negative log likelihood),

$$L(y, \hat{y}) = f_{CEL}(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{C} y_{n,i} \log\left(\hat{y}_{n,i}\right)$$

Cross-Entropy Loss (also known as negative log likelihood),

$$L(y, \hat{y}) = f_{CEL}(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{C} y_{n,i} \log\left(\hat{y}_{n,i}\right)$$

Stochastic Gradient Descent

with *momentum*

and *learning rate scheduler*.

Stochastic Gradient Descent
with *momentum*
and *learning rate scheduler*.

Operation Tests

ReLU vs Sigmoid vs f_{fit}

Operation Tests
ReLU vs Sigmoid vs f_{fit}

Conclusion

Conclusion

Overview

I assembled an experimental setup from scratch

I characterized the response of the MRR in several aspects

I implemented the bistable response in standard software libraries

Overview

I assembled an experimental setup from scratch

I characterized the response of the MRR in several aspects

I implemented the bistable response in standard software libraries

Future Perspectives

Continue the current work with a quantitative analysis of specific features

Future Perspectives

Continue the current work with a quantitative analysis of specific features

Future Perspectives

Continue the current work with a quantitative analysis of specific features

Enhance the physical theory to describe time dependent phenomena

Future Perspectives

Continue the current work with a quantitative analysis of specific features

Enhance the physical theory to describe time dependent phenomena

Future Perspectives

Continue the current work with a quantitative analysis of specific features

Enhance the physical theory to describe time dependent phenomena

Proceed with the development of the simulations to include all the characteristics of the physical system

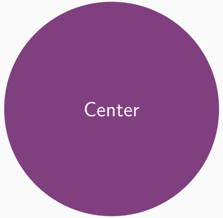
Future Perspectives

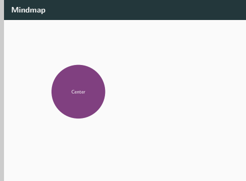
Continue the current work with a quantitative analysis of specific features

Enhance the physical theory to describe time dependent phenomena

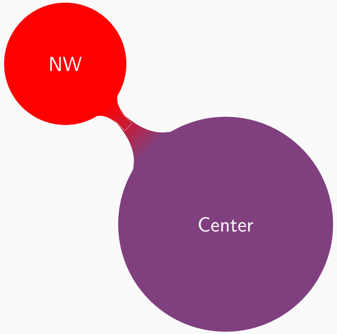
Proceed with the development of the simulations to include all the characteristics of the physical system

Mindmap

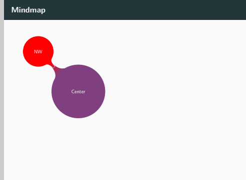




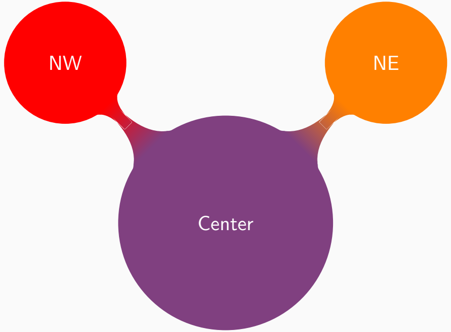
Mindmap



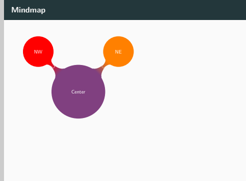
2018-03-09



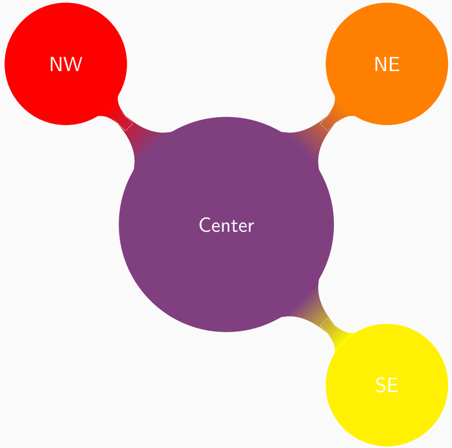
Mindmap



2018-03-09

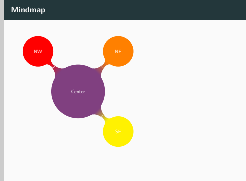


Mindmap

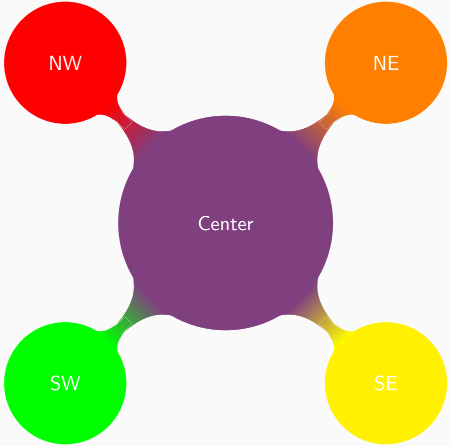


2018-03-09

All-Optical Neural Networks
└─Conclusion
└─Mindmap



Mindmap



All-Optical Neural Networks
└─Conclusion

2018-03-09

└─Mindmap