Breeding Artificial Brains on Supercomputers

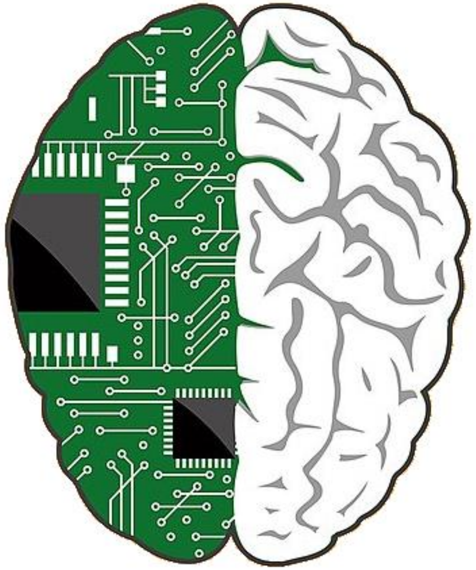
MD HPC UG – May 15th 2019

Aaron Vose

avose@cray.com

Breeding NNs with HPC – Presentation Overview

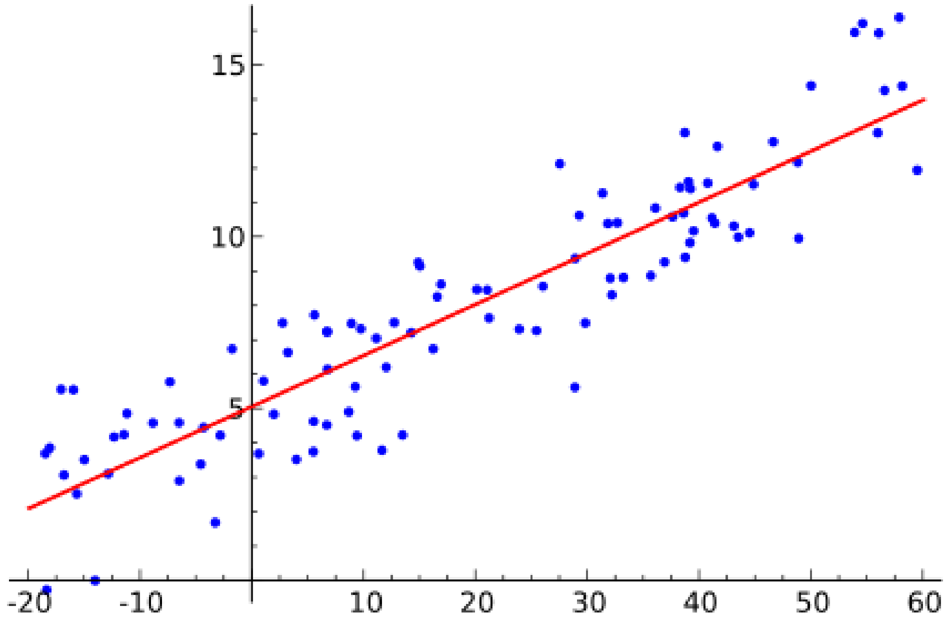
- Introduction and motivation



https://en.wikipedia.org/wiki/File:Carrots_of_many_colors.jpg
https://en.wikipedia.org/wiki/File:RightBrainMicroprocessor.jpg

Breeding NNs with HPC – Presentation Overview

- Introduction and motivation
- High-level overview of artificial neural networks
  - Perceptron as a model of a neuron
  - Connected perceptrons as a model of a brain
  - Backpropagation as a model of learning



$$y = mx + b$$

https://upload.wikimedia.org/wikipedia/commons/3/3a/Linear_regression.svg

Breeding NNs with HPC – Presentation Overview

- Introduction and motivation
- High-level overview of artificial neural networks
  - Perceptron as a model of a neuron
  - Connected perceptrons as a model of a brain
  - Backpropagation as a model of learning

?

Choose Hyperparameters

https://www.flickr.com/photos/bfishadow/4407860229

Breeding NNs with HPC – Presentation Overview

- Introduction and motivation

- High-level overview of artificial neural networks
  - Perceptron as a model of a neuron
  - Connected perceptrons as a model of a brain
  - Backpropagation as a model of learning

- Explanation of hyperparameter optimization
  - grid search, random search, **genetic algorithms** ?
  - Bayesian approaches, reinforcement learning, etc.

Breeding NNs with HPC – Presentation Overview

- Introduction and motivation

- High-level overview of artificial neural networks
  - Perceptron as a model of a neuron
  - Connected perceptrons as a model of a brain
  - Backpropagation as a model of learning

Artificial Brains

- Explanation of hyperparameter optimization
  - grid search, random search, **genetic algorithms** → **Breeding**
  - Bayesian approaches, reinforcement learning, etc.

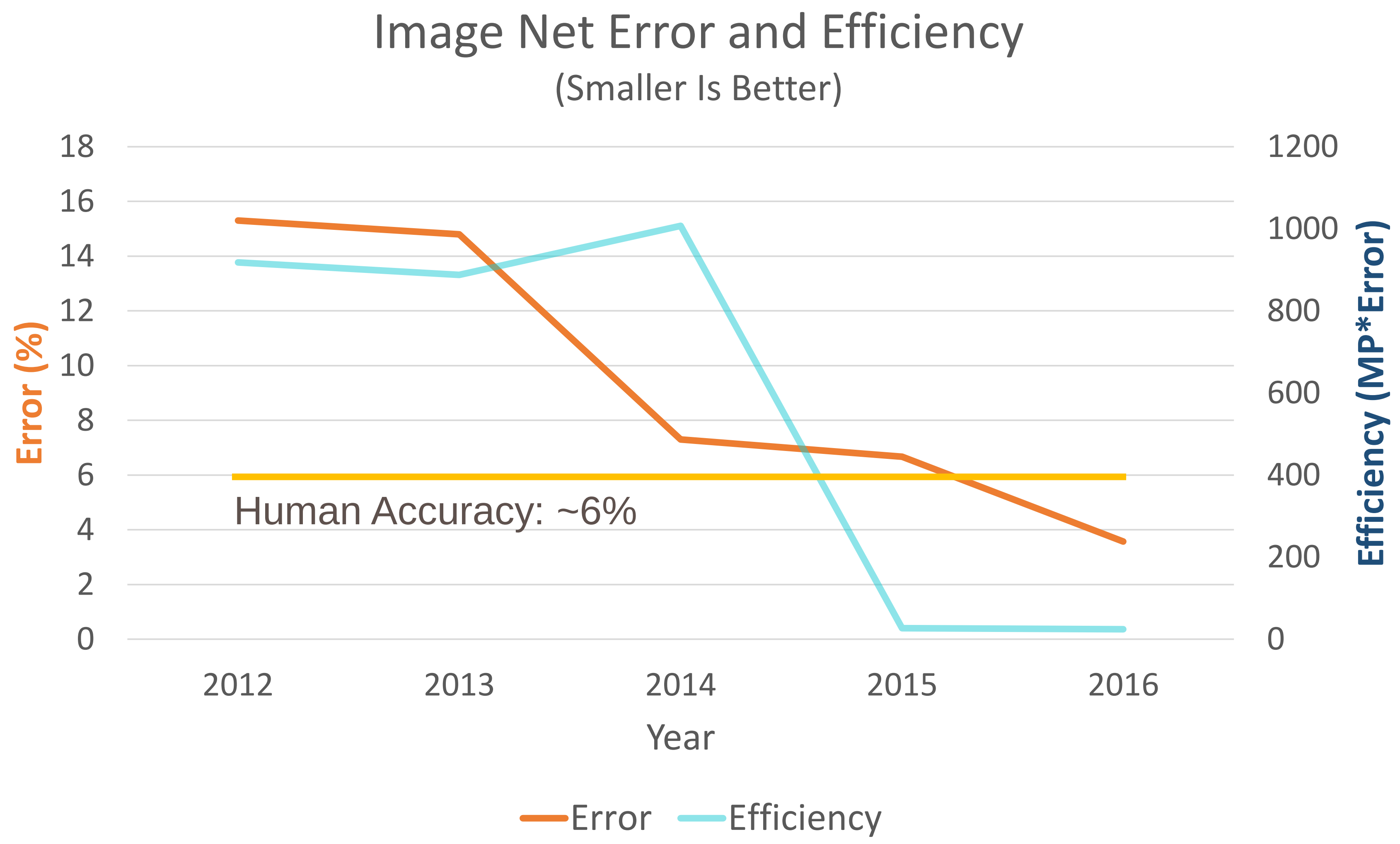
“

THE FUTURE IS SELDOM THE SAME AS THE PAST.

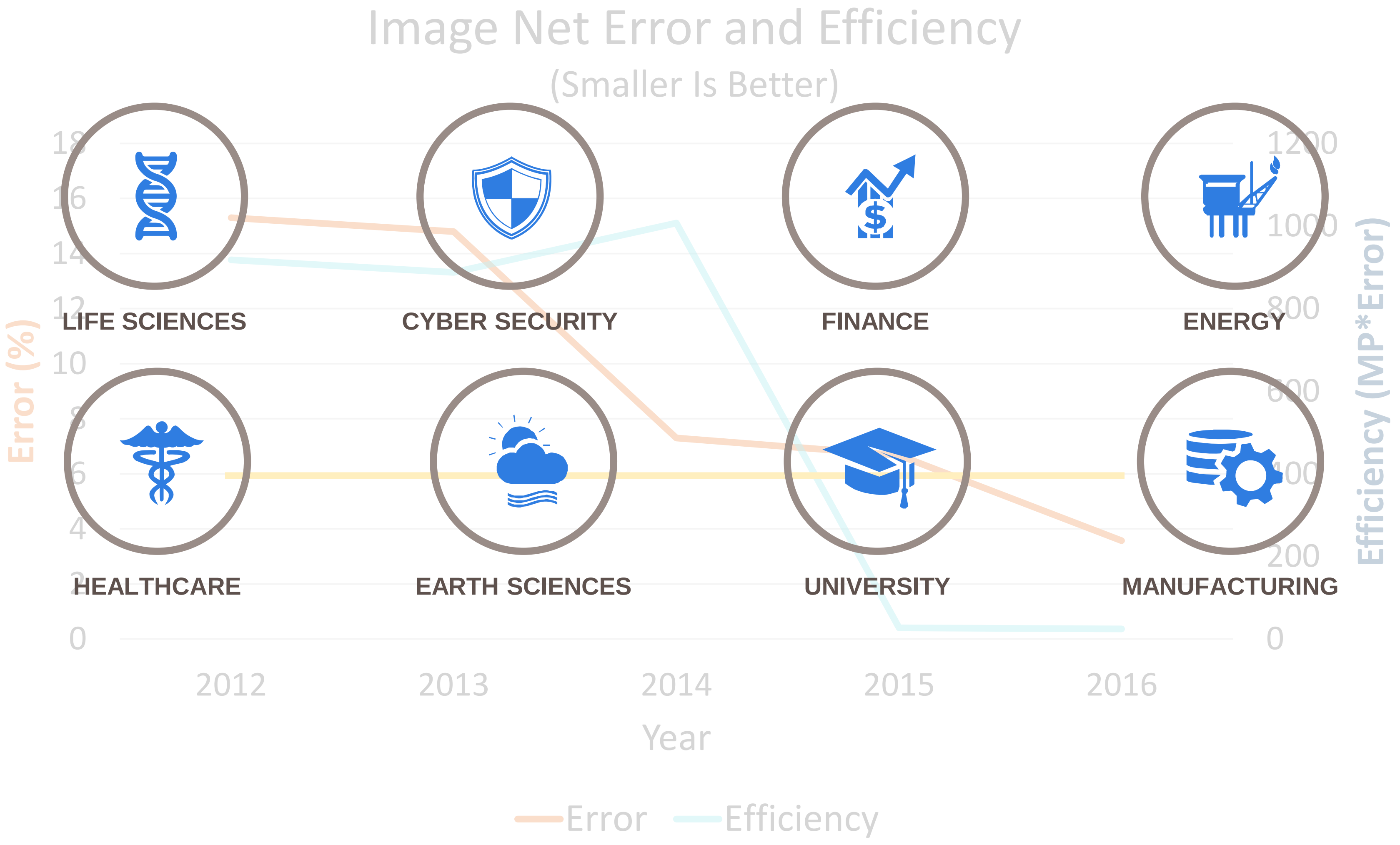
”

SEYMOUR CRAY

Breeding NNs with HPC – Superhuman

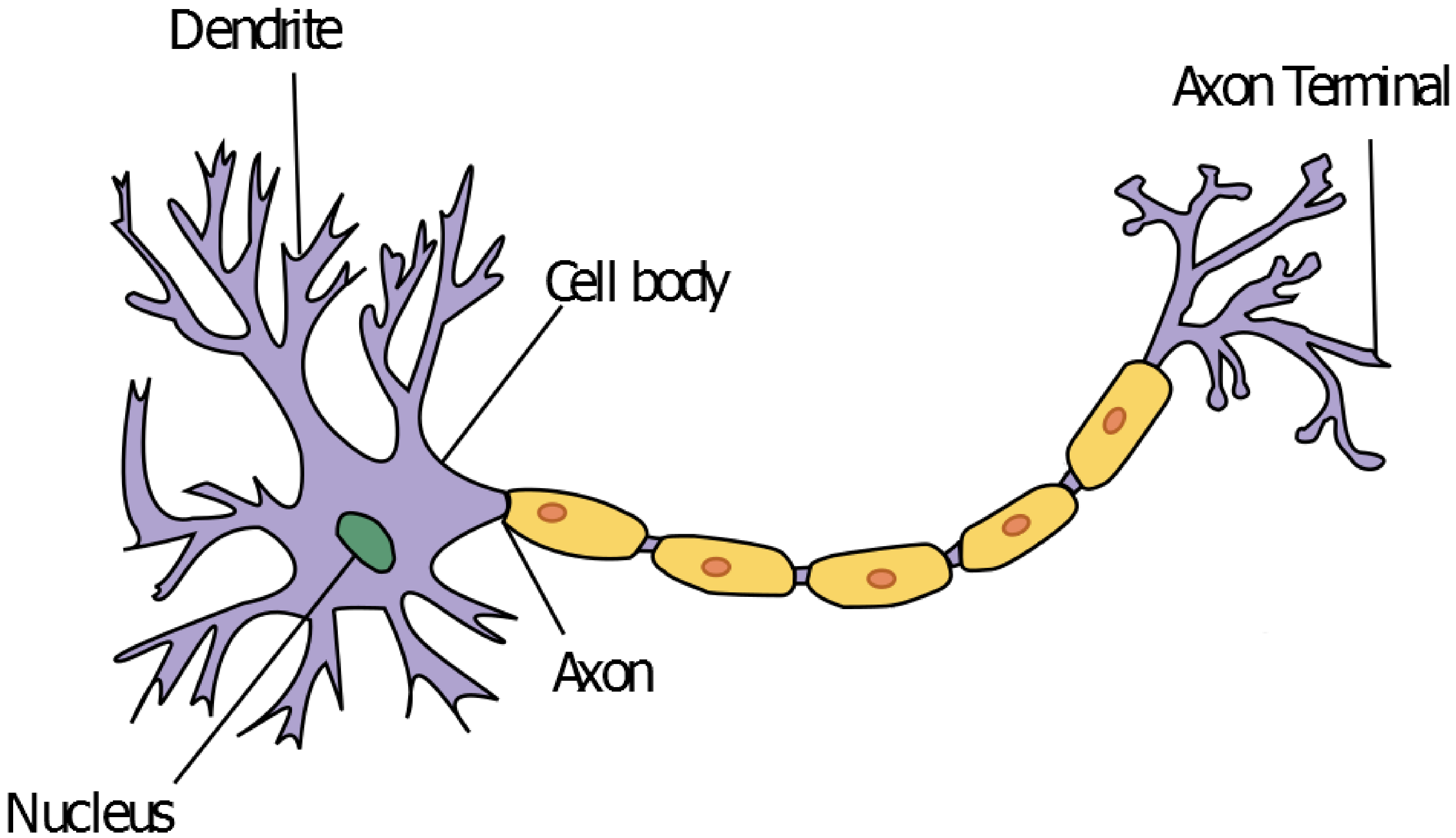


Breeding NNs with HPC – Cross-Industry



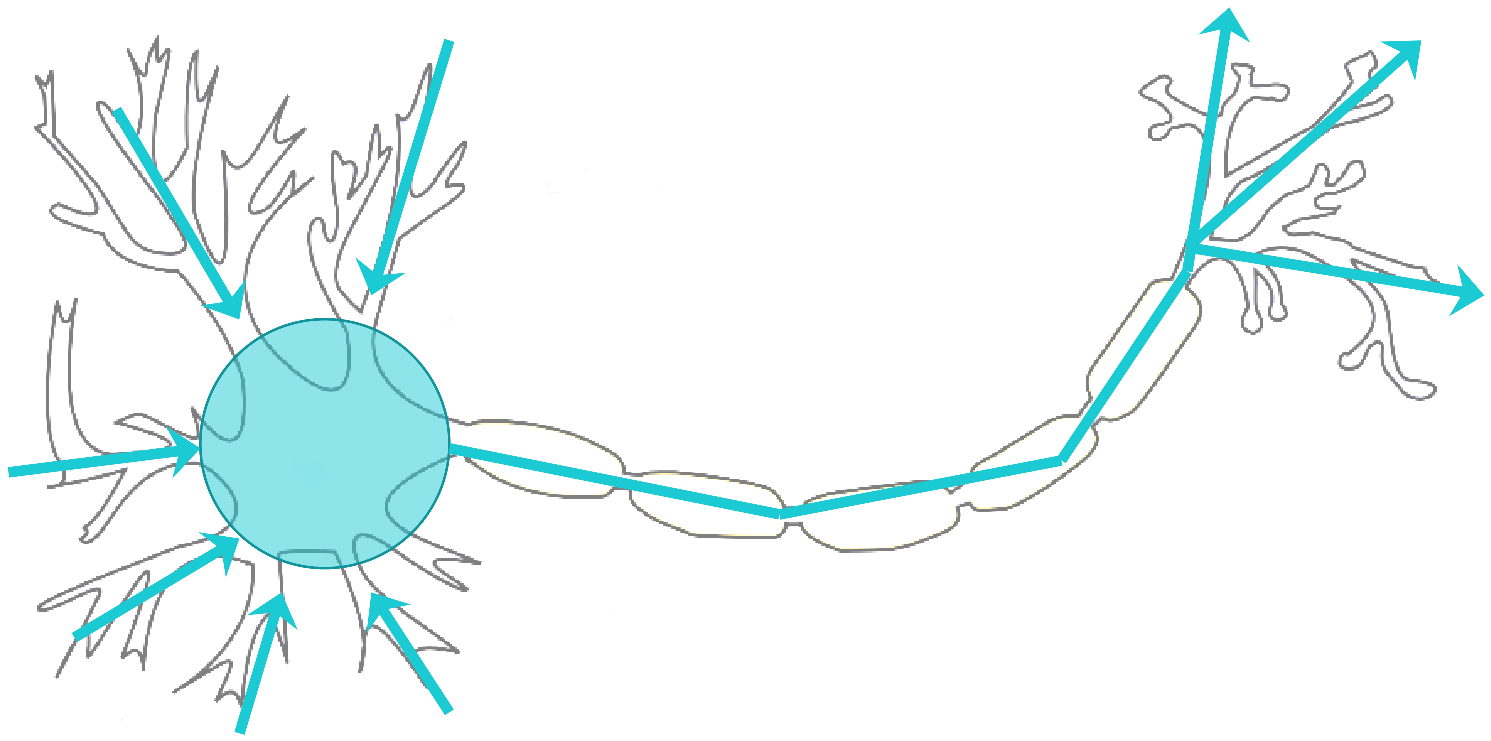
Neural Network Introduction

ANN Overview – Perceptron Model – Neuron



- Perceptron as a model for a neuron (Frank Rosenblatt, 1957)

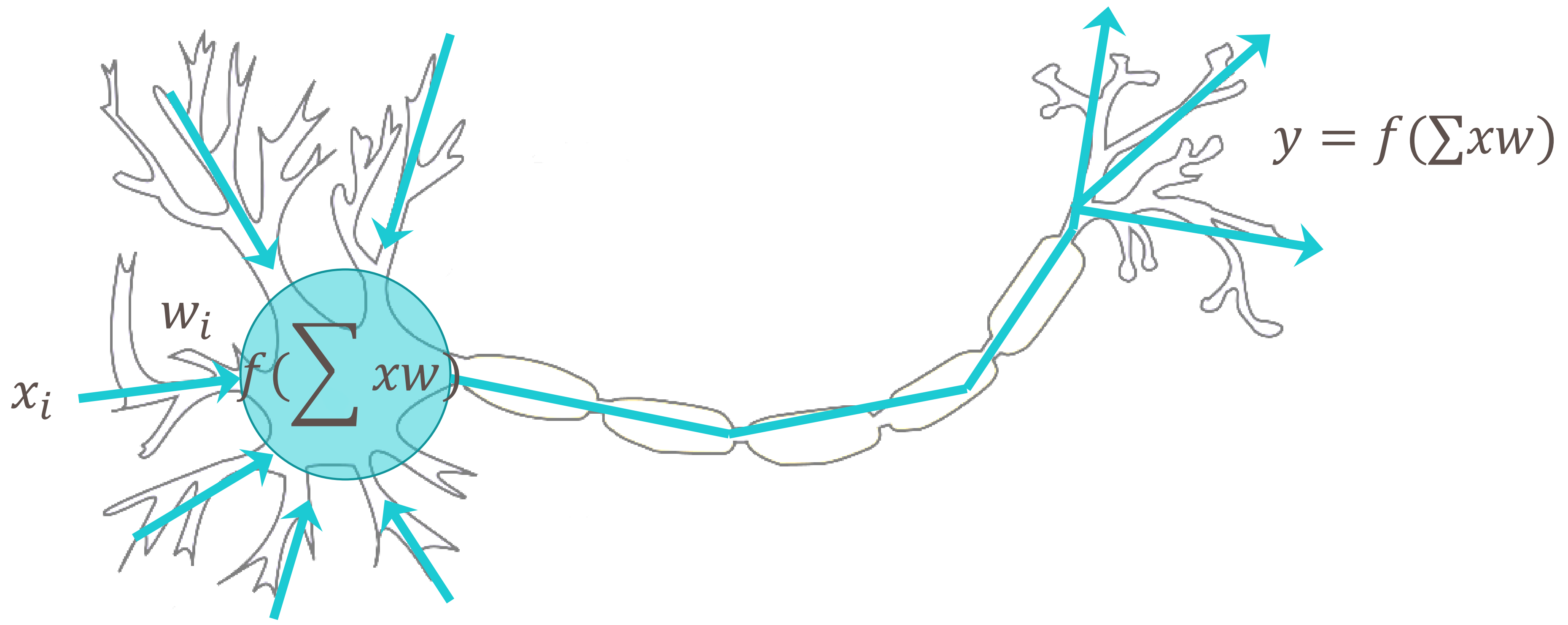
ANN Overview – Perceptron Model – Info Flow



- Information flows in dendrites, is processed in the nucleus, flows out axon.

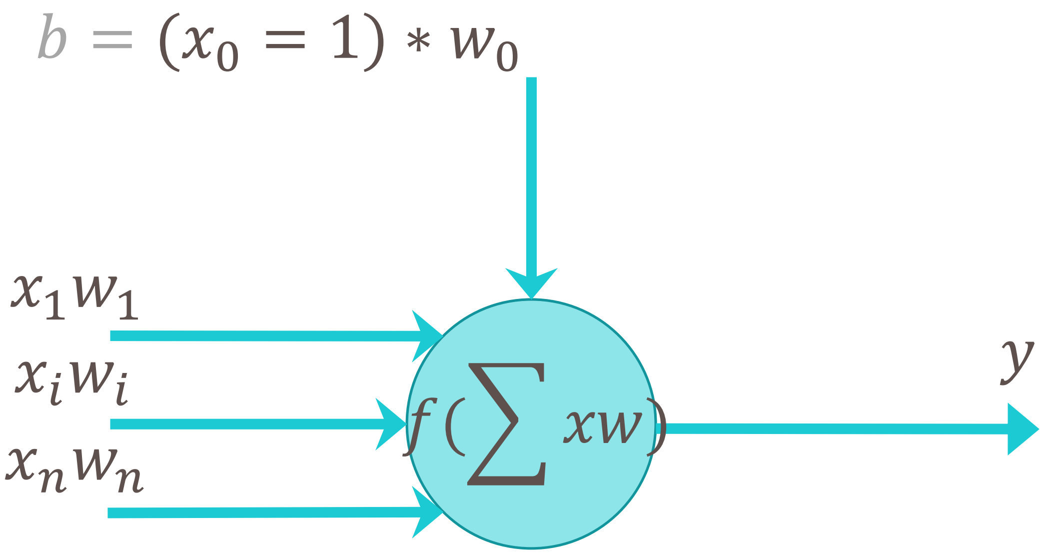
https://commons.wikimedia.org/wiki/File:Neuron.svg

ANN Overview – Perceptron Model – Info Flow

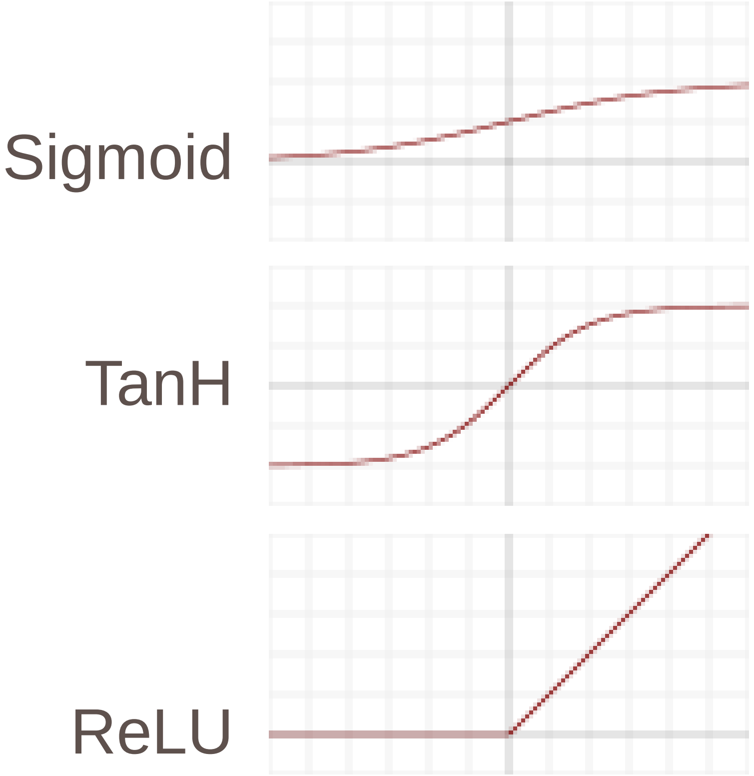


- Information flows in dendrites, is processed in the nucleus, flows out axon.

ANN Overview – Perceptron Model – Linear Model

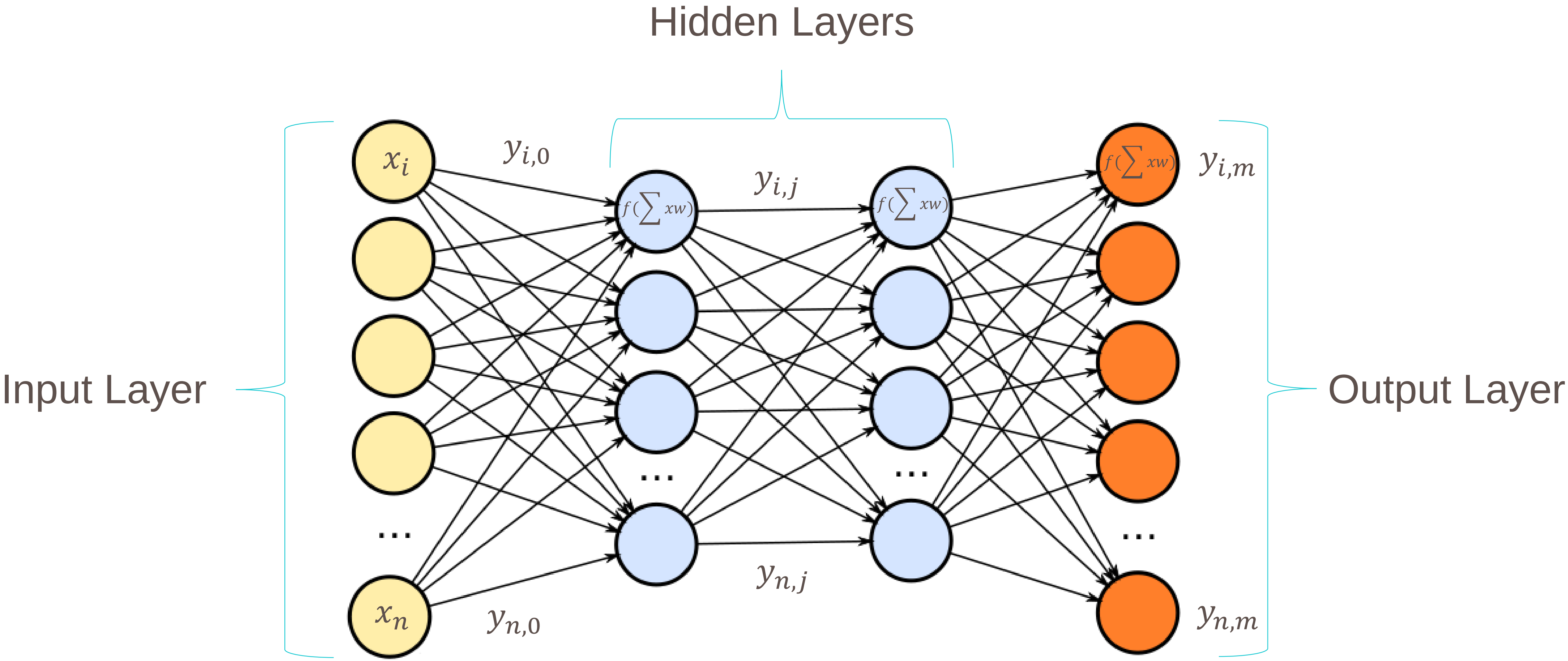


$y = f(\sum xw)$



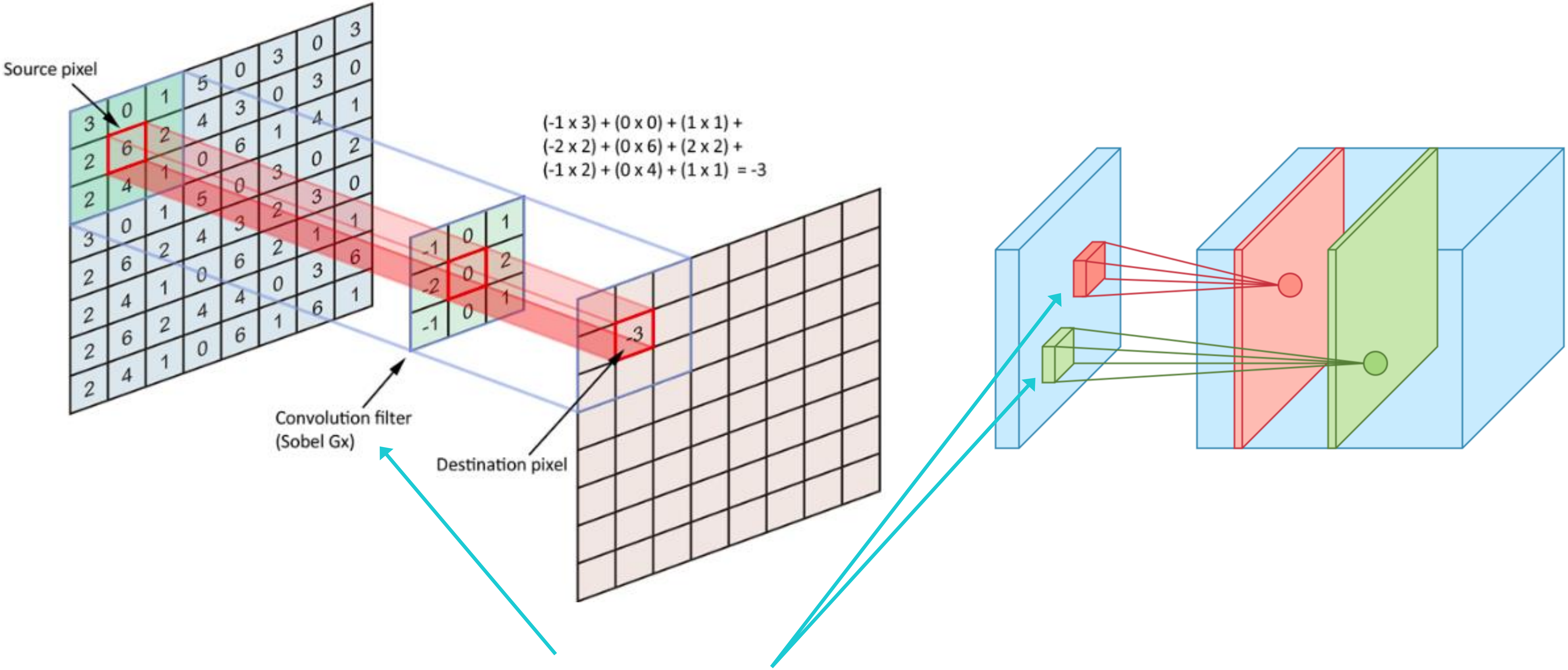
- Linear model wrapped by a non-linear function:
  - $\boldsymbol{y = f(wx + b)}$

ANN Overview – Deep Neural Networks



https://commons.wikimedia.org/wiki/File:Neural_network_bottleneck_achitecture.svg

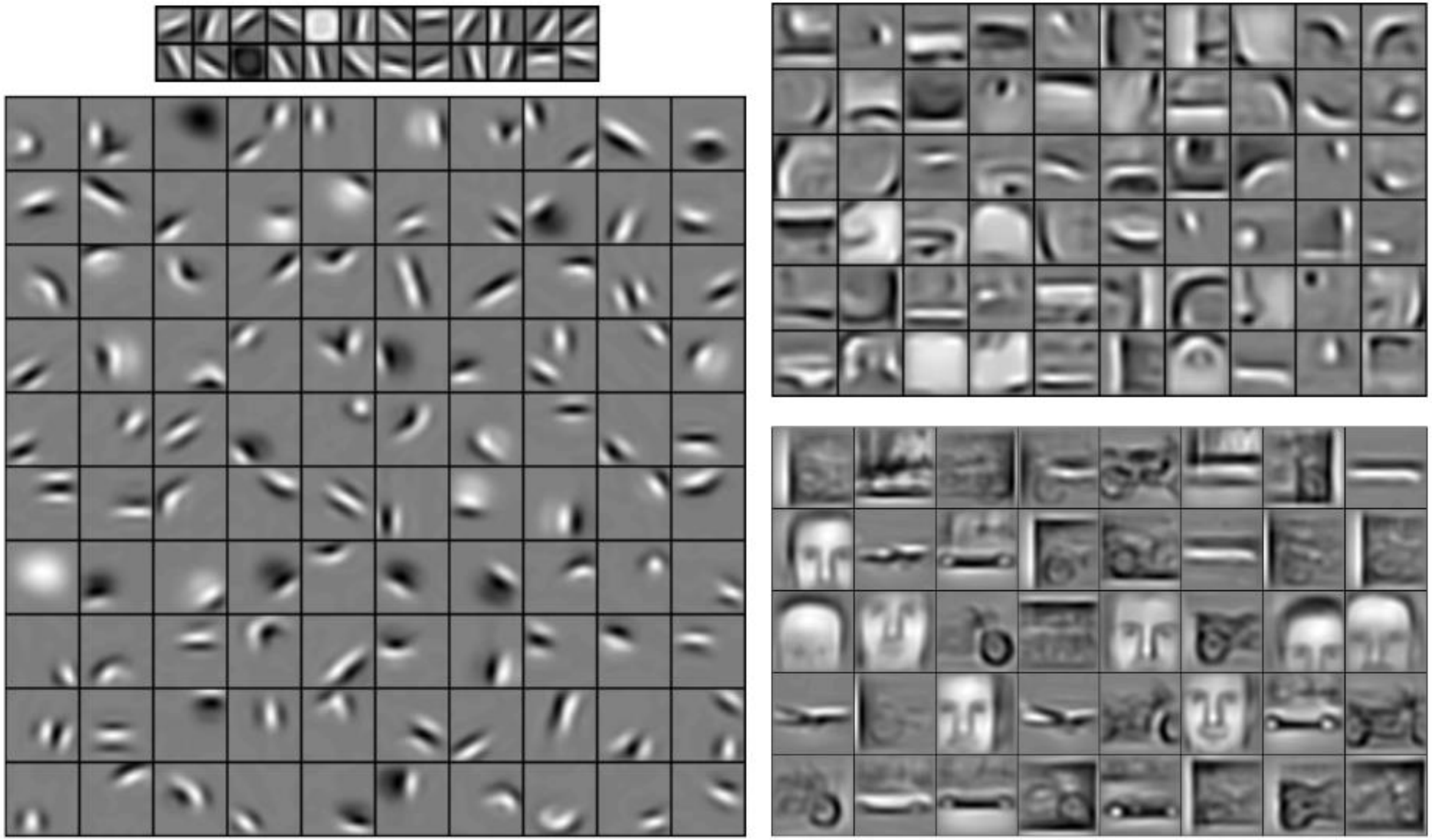
ANN Overview – Convolutional Neural Networks



weights w_i

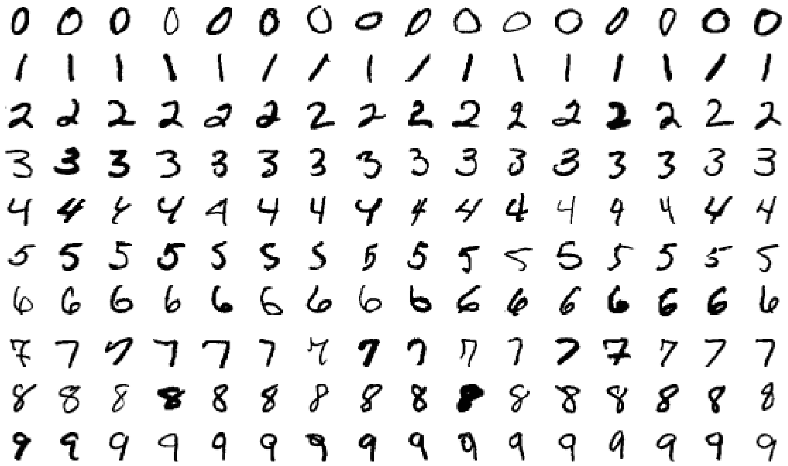
https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

ANN Overview – CNN Filters Visualized



http://web.eecs.umich.edu/~honglak/icml09-ConvolutionalDeepBeliefNetworks.pdf

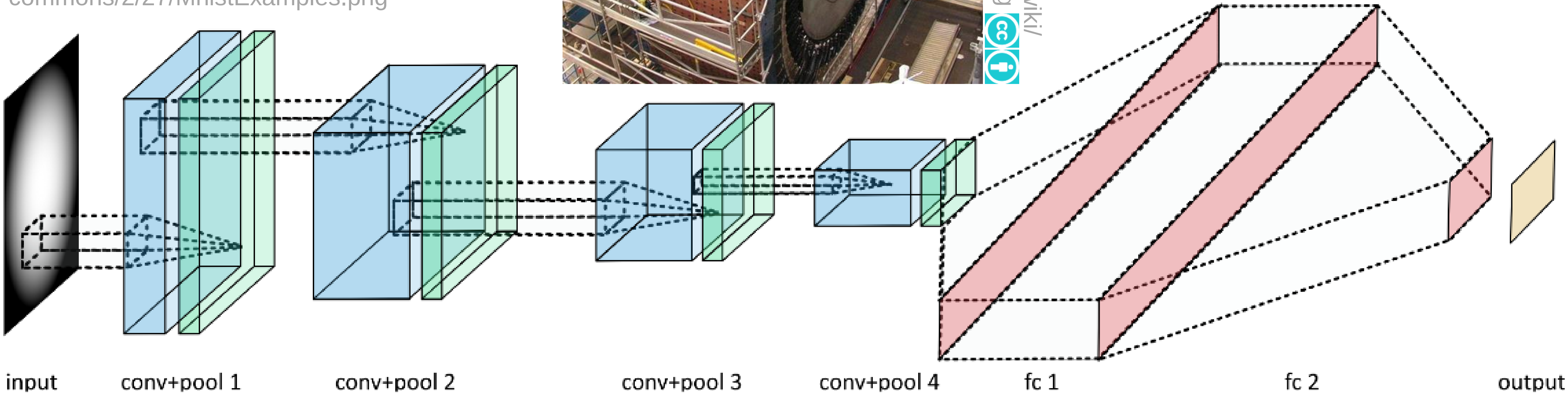
ANN Overview – CNN Architecture



https://upload.wikimedia.org/wikipedia/commons/2/27/MnistExamples.png

https://commons.wikimedia.org/wiki/File:ATLAS_Tile_Calorimeter.png

https://www.researchgate.net/publication/221362351_Modeling_Pixel_Means_and_Covariances_Using_Factorized_Third-Order_Boltzmann_Machines



https://arxiv.org/abs/1711.03573

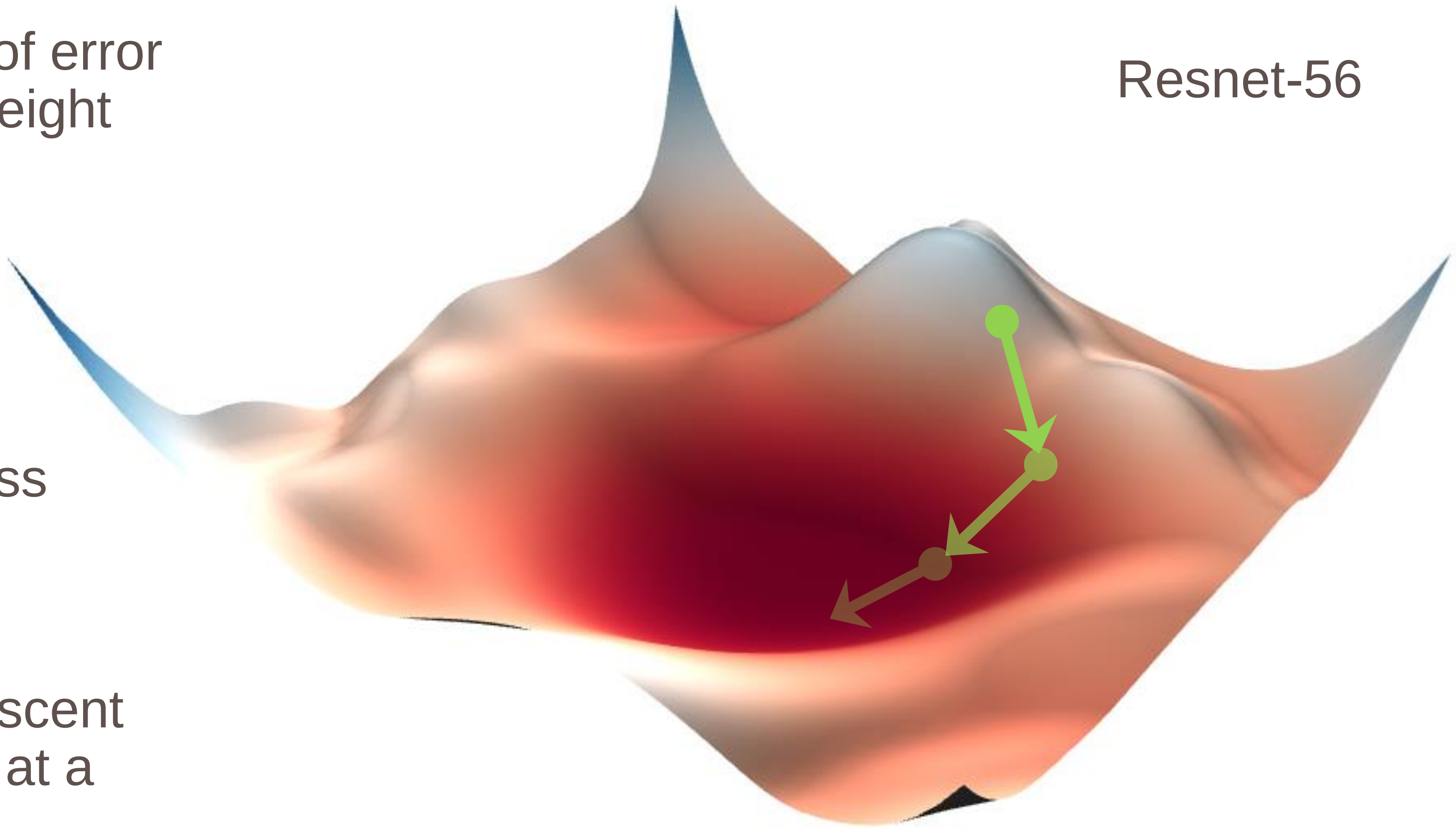
ANN Overview – Backpropagation

- Find partial derivative of error with respect to each weight (chain-rule):

$$\nabla w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

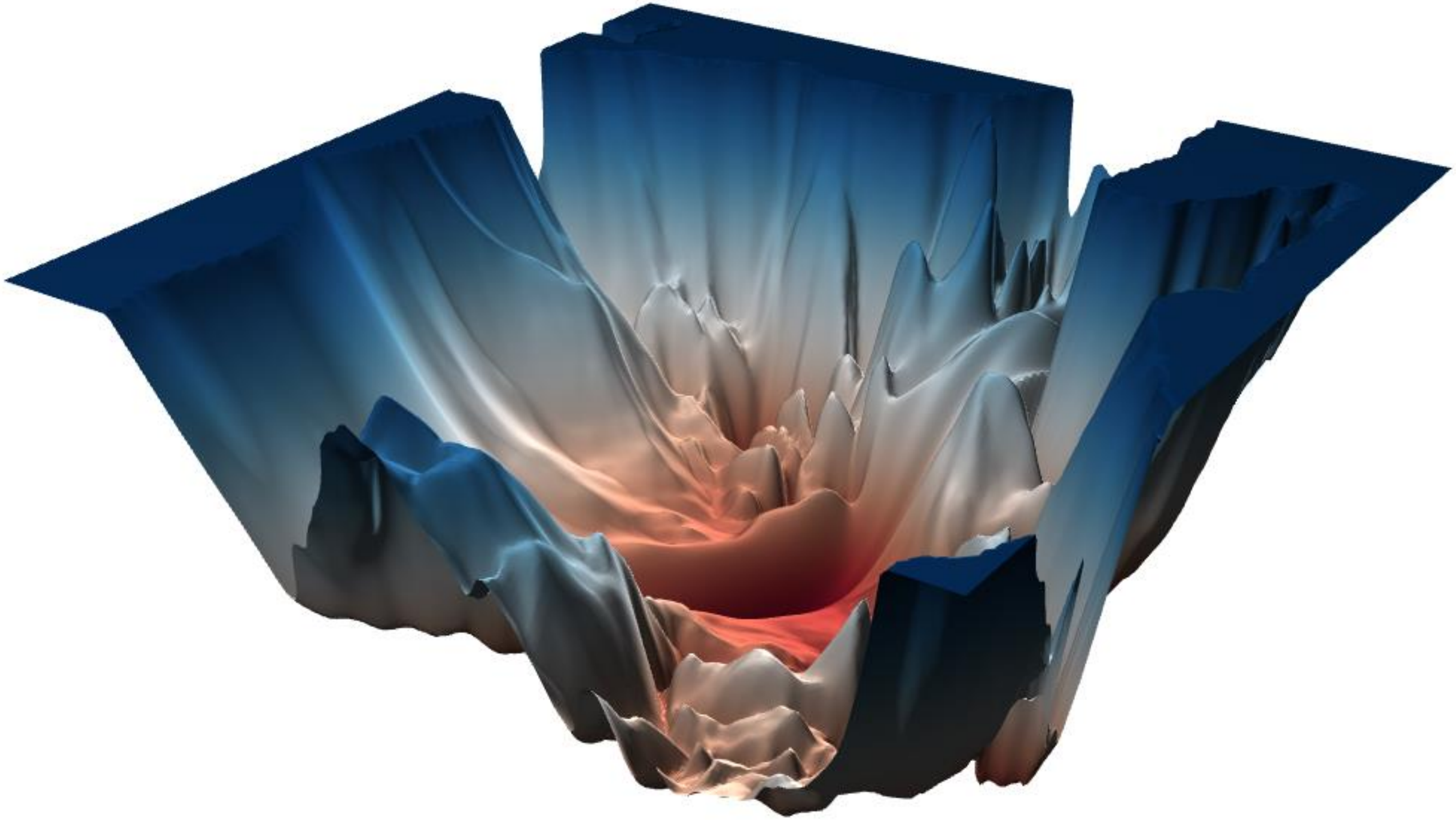
- Challenge to compute, synchronize ∇w_{ij} across compute elements

- Stochastic gradient descent (SGD) with mini-batch at a time



http://www.telesens.co/loss-landscape-viz/viewer.html
https://arxiv.org/pdf/1712.09913.pdf

ANN Overview – Parameter Optimization



Resnet-56 (no skip)

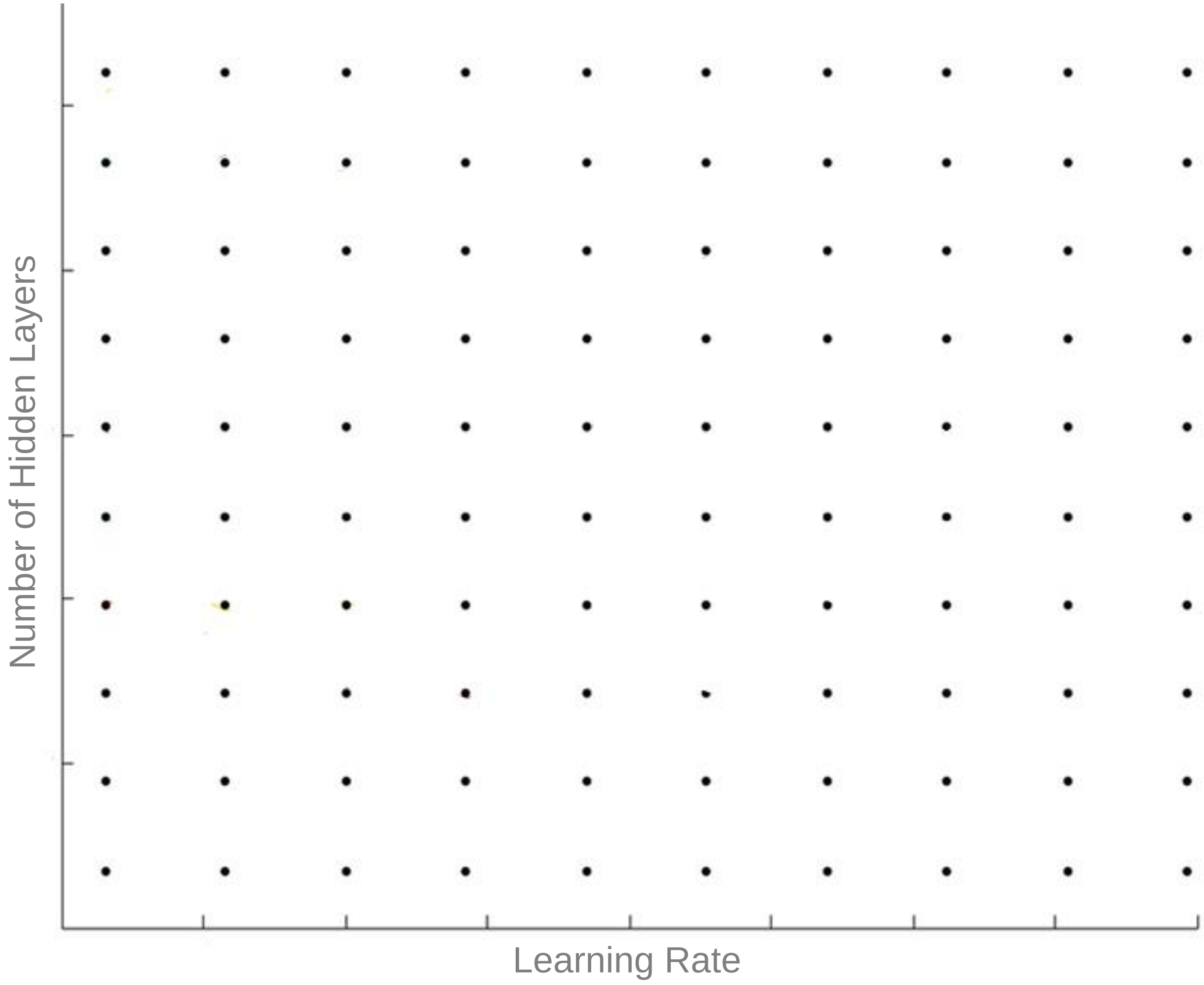
http://www.telesens.co/loss-landscape-viz/viewer.html
https://arxiv.org/pdf/1712.09913.pdf

Hyperparameter Optimization

NN HPO – Basic Grid Search

- Simple
- Easily parallelizable

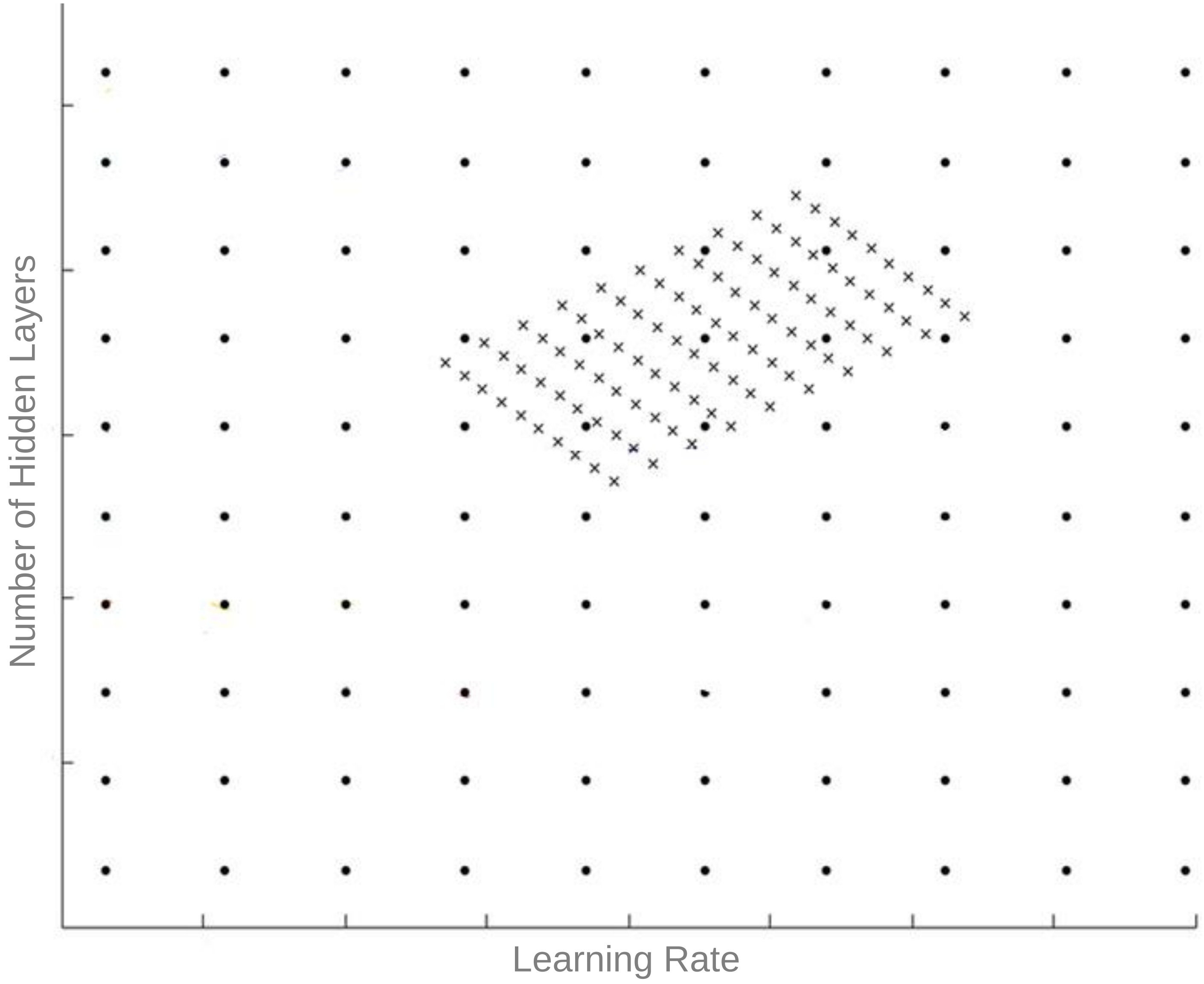
- Curse of dimensionality
- Computation expense



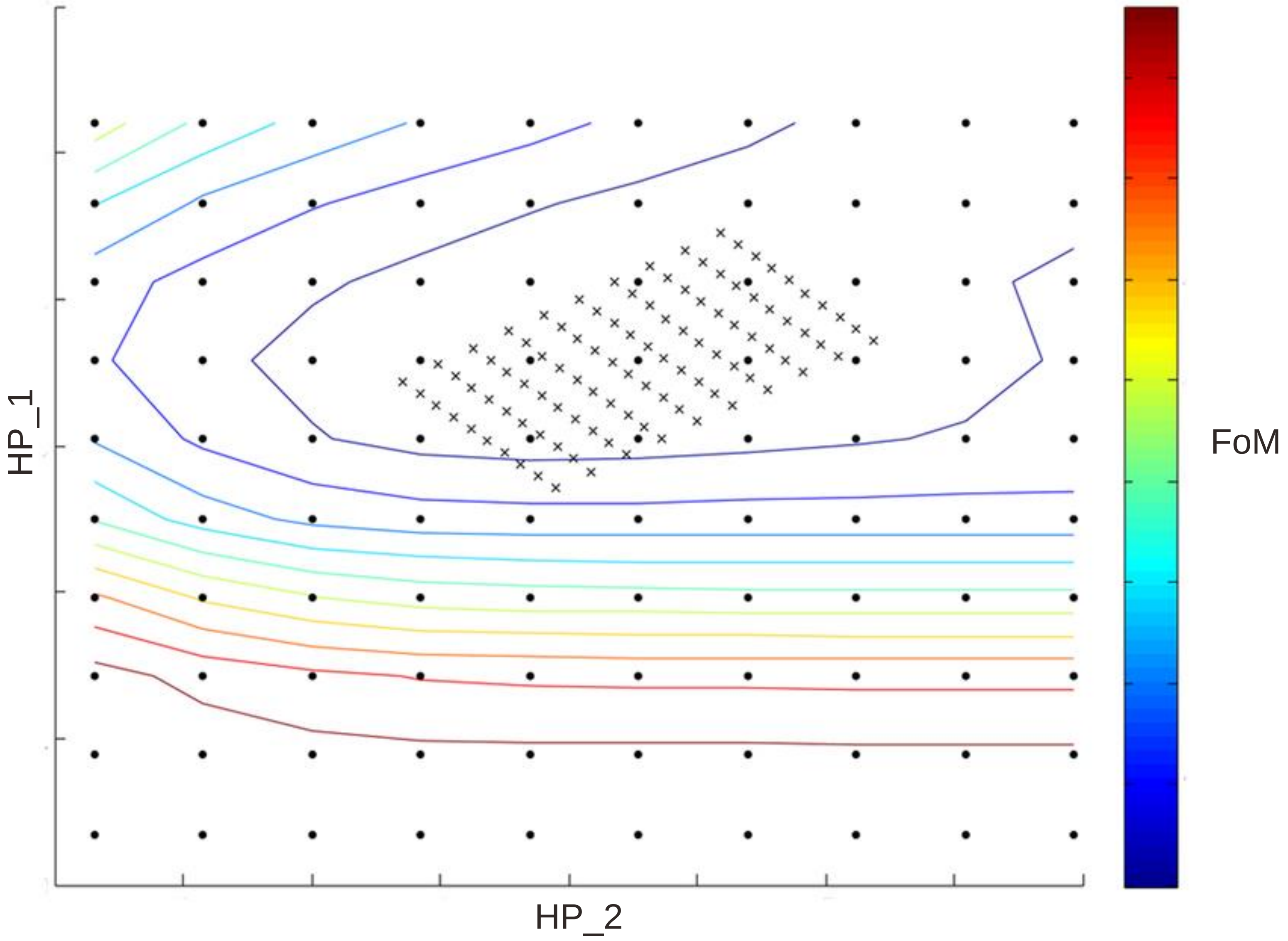
NN HPO – Iterative Grid Search

- Simple
- Easily parallelizable

- Curse of dimensionality
- Computation expense



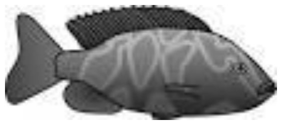
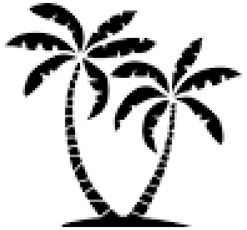
NN HPO – Iterative Grid Search with FoM Surface



NN HPO – Genetic Algorithms

- Think of a GA on HPO as:
  - “Automatic, iterative, stochastic grid search with pruning”

- Inspired by biological systems found in nature:
  - Mutation
  - Crossover
  - Selection


NN HPO – Genetic Algorithm Generation Cycle

Founder

Mutation

Create initial population around default HPs.

Train and get NNs' accuracy.

Evaluate

Selection

Choose pair of parent NNs.

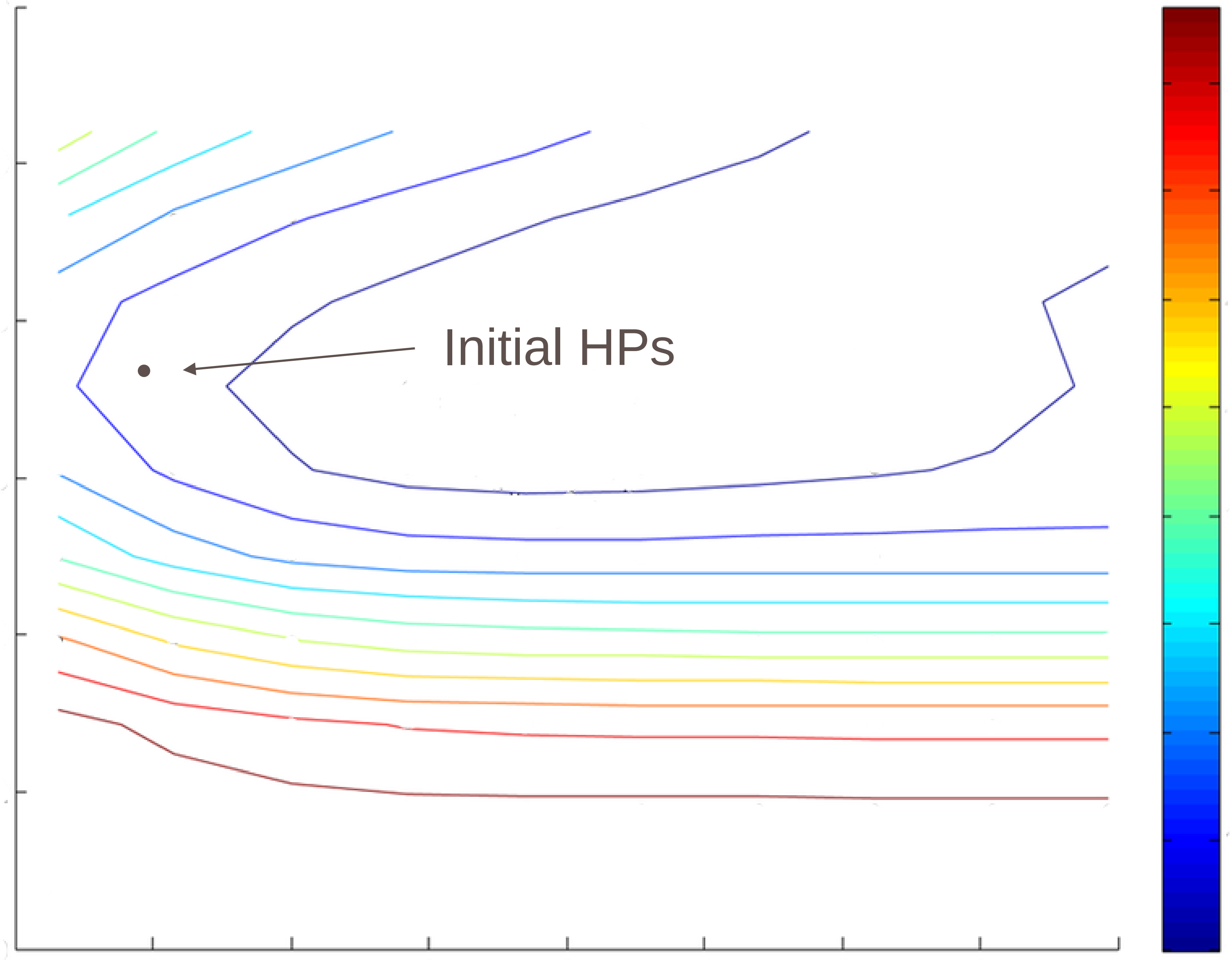
Reproduce

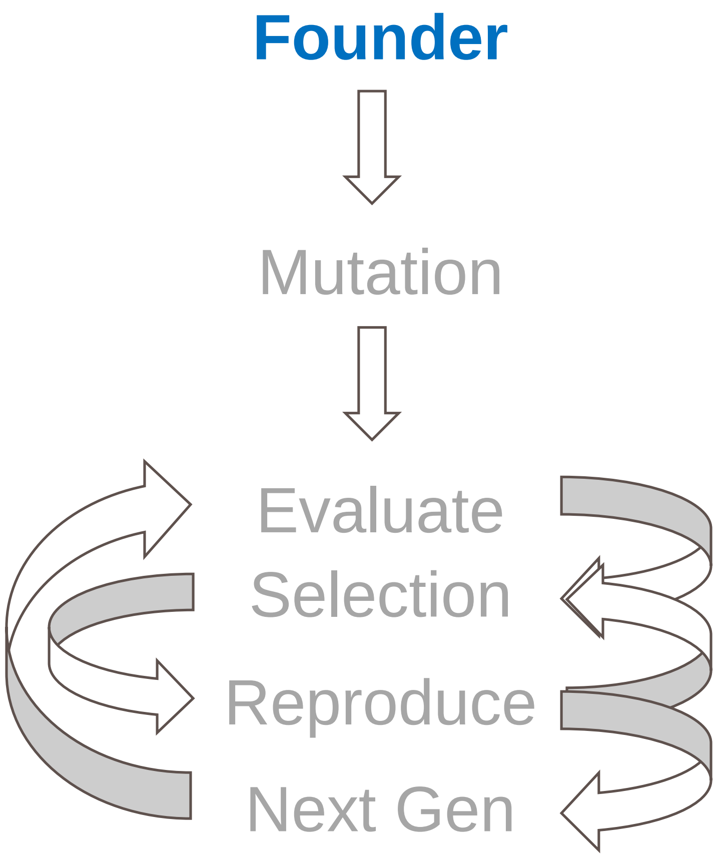
Create child from parents with crossover and mutation.

Next Gen

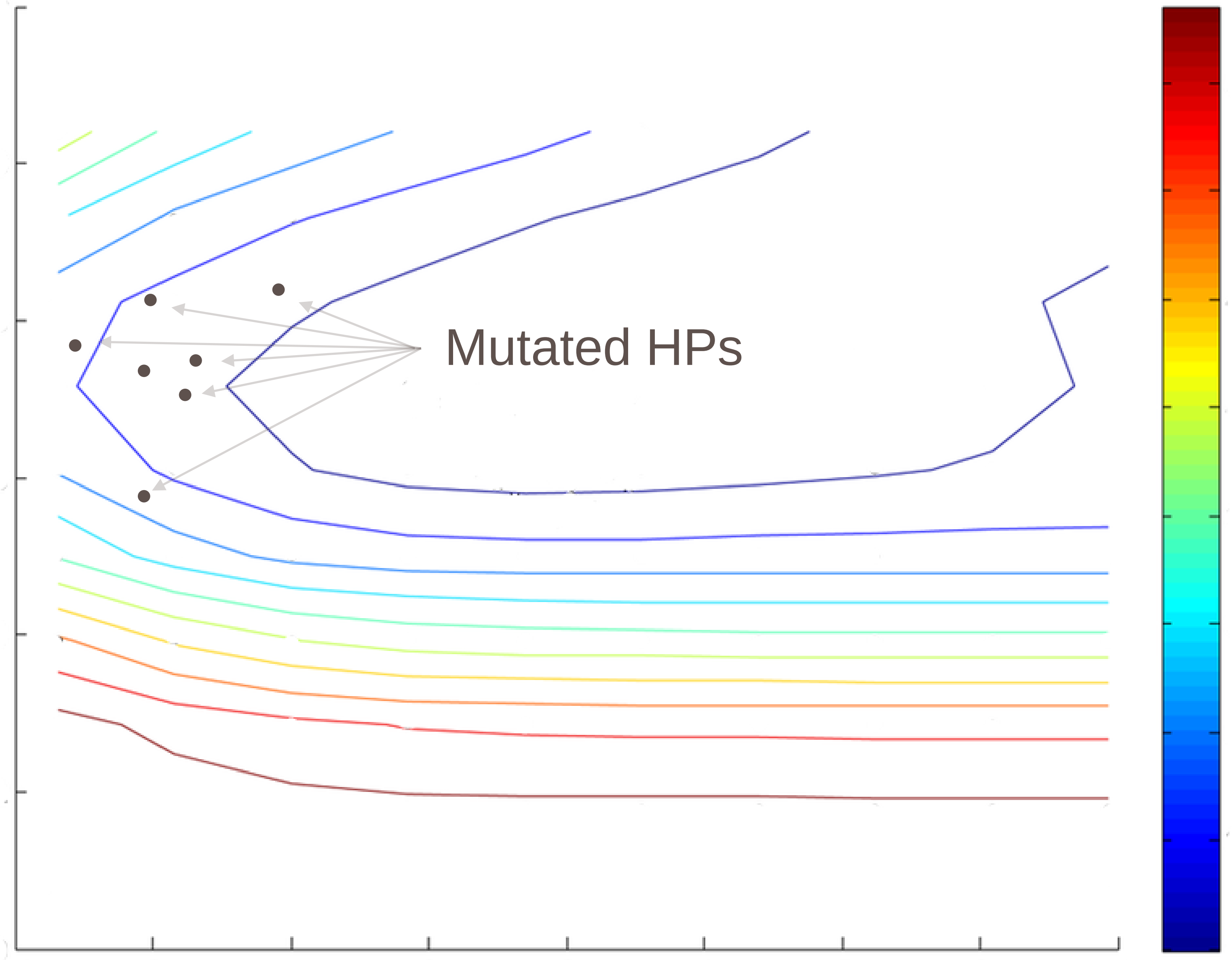
Kill old population; children "grow up".

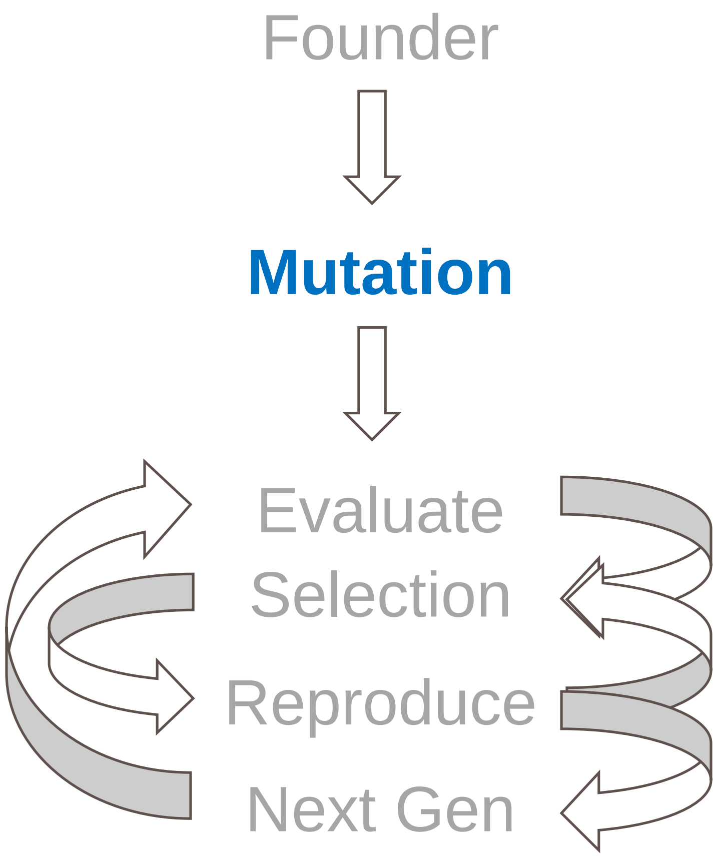
NN HPO – HPO with GAs: Founder



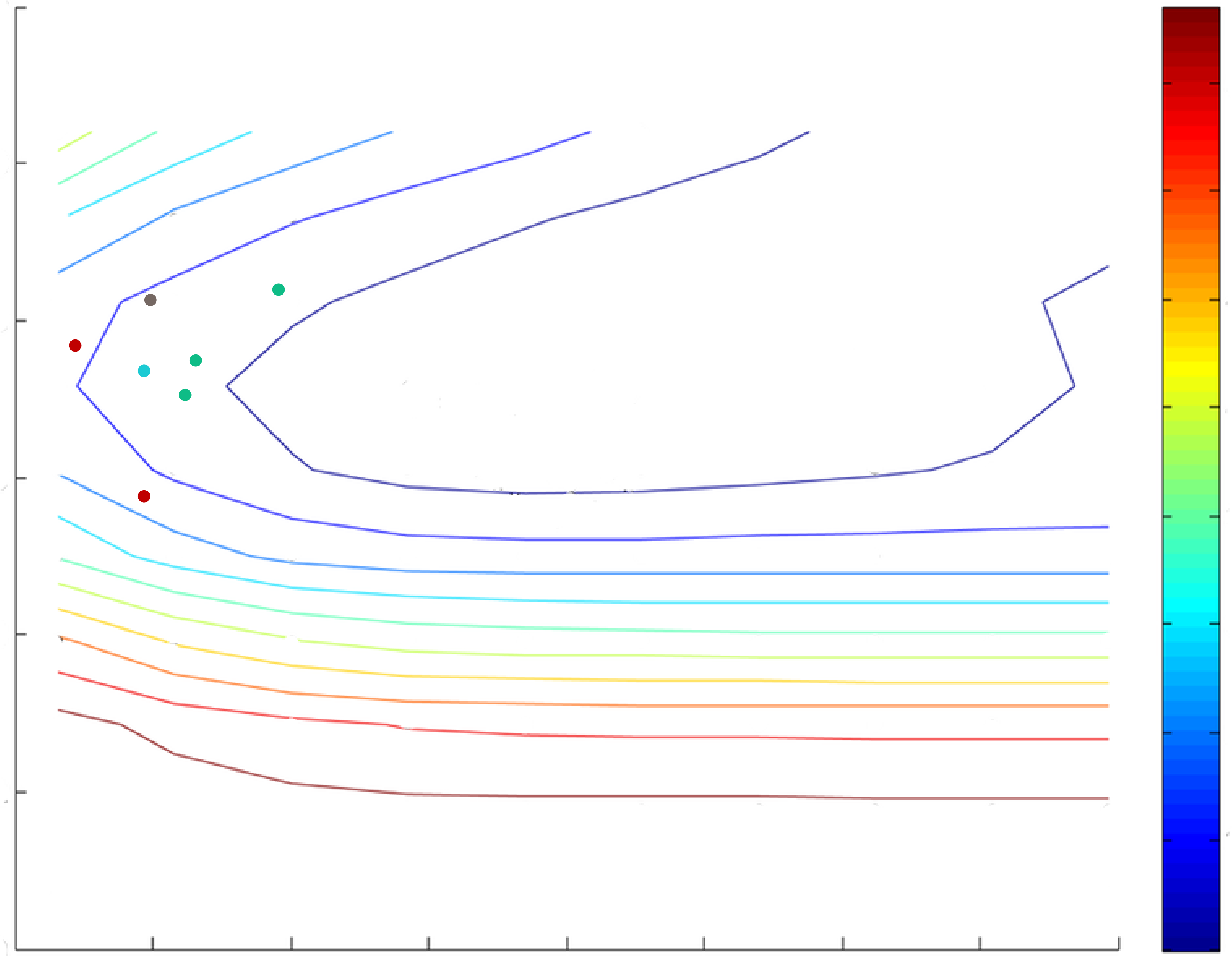


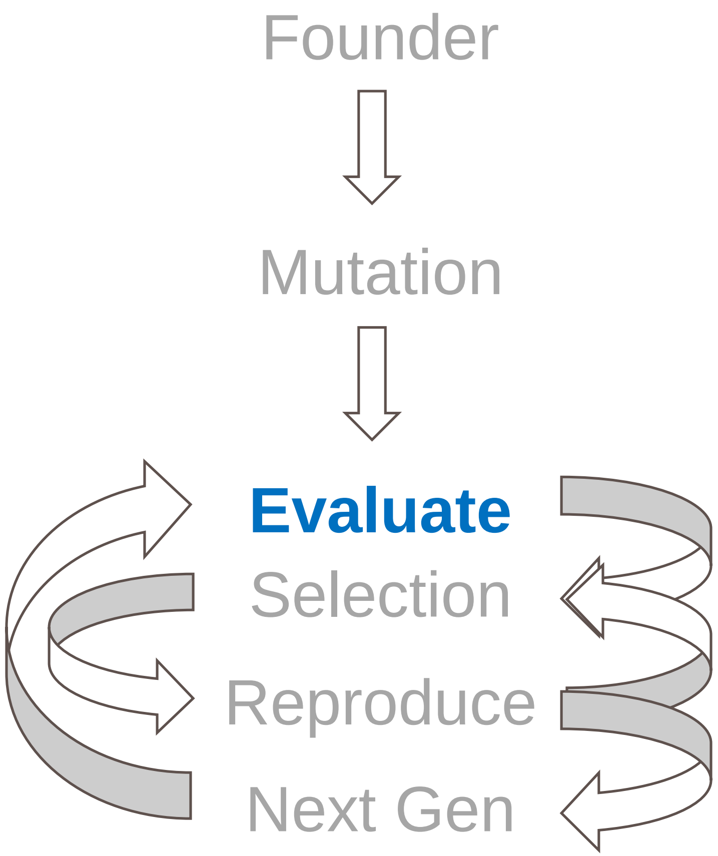
NN HPO – HPO with GAs: Initial Mutation



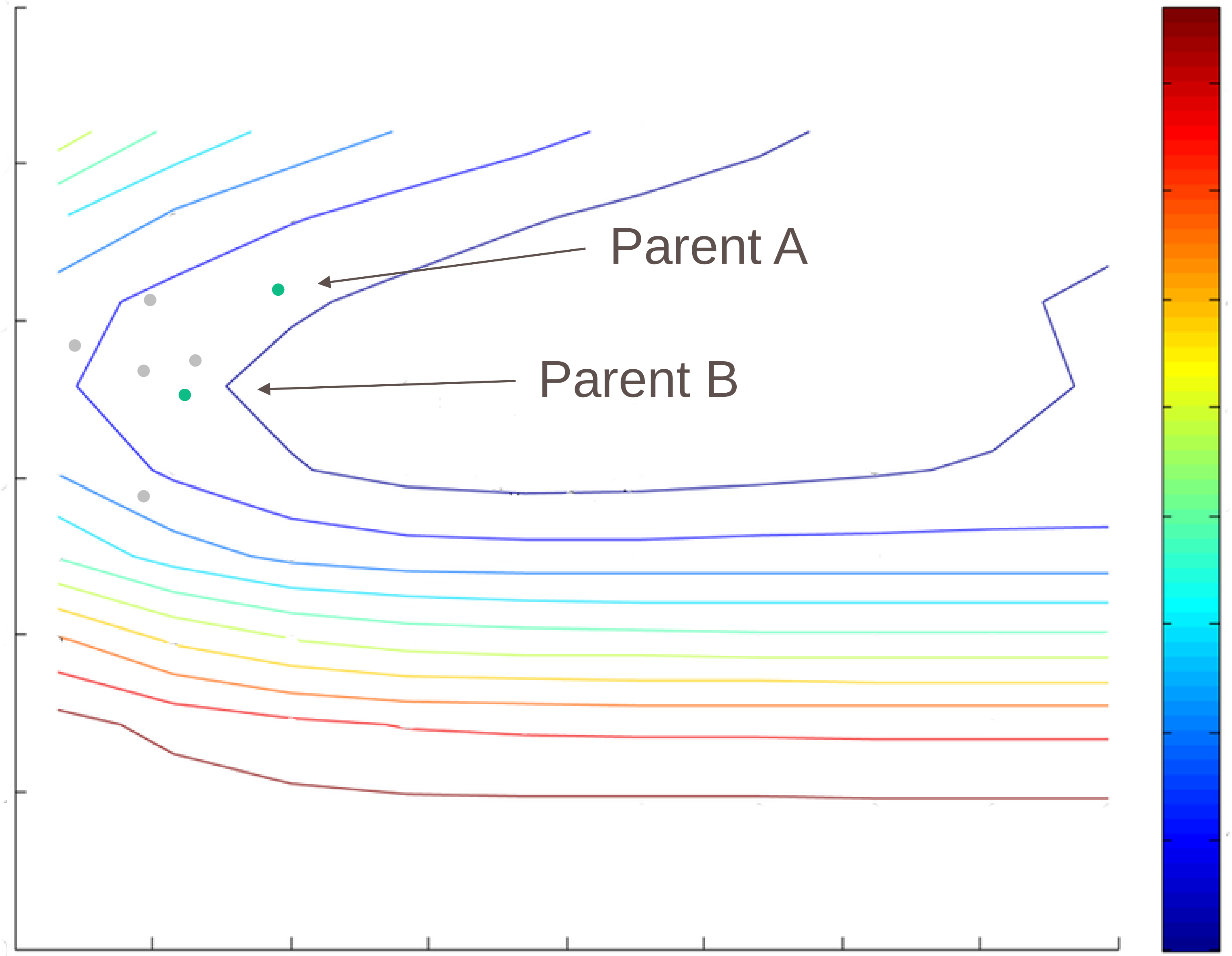


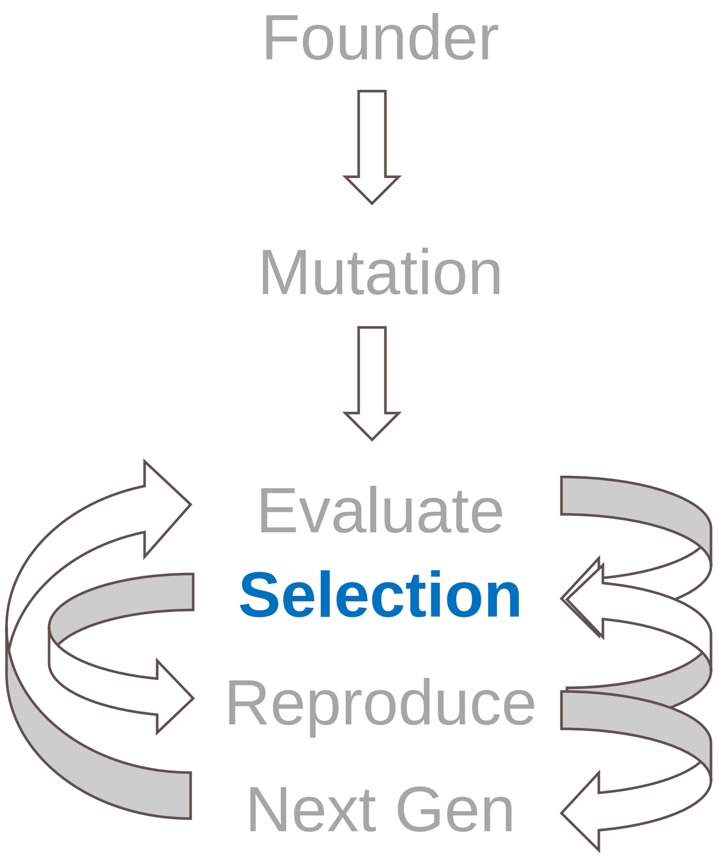
NN HPO – HPO with GAs: Evaluate Fitness 1



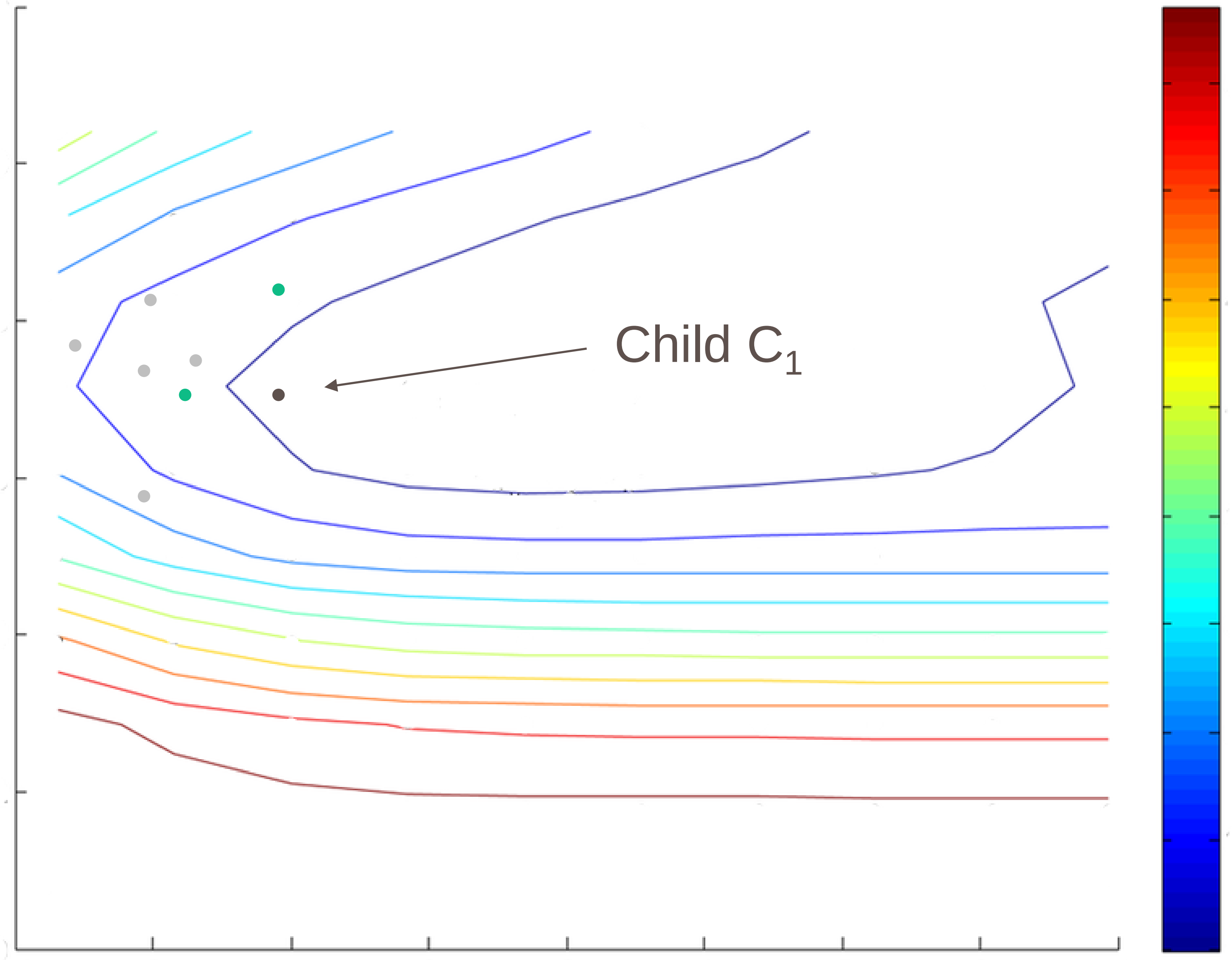


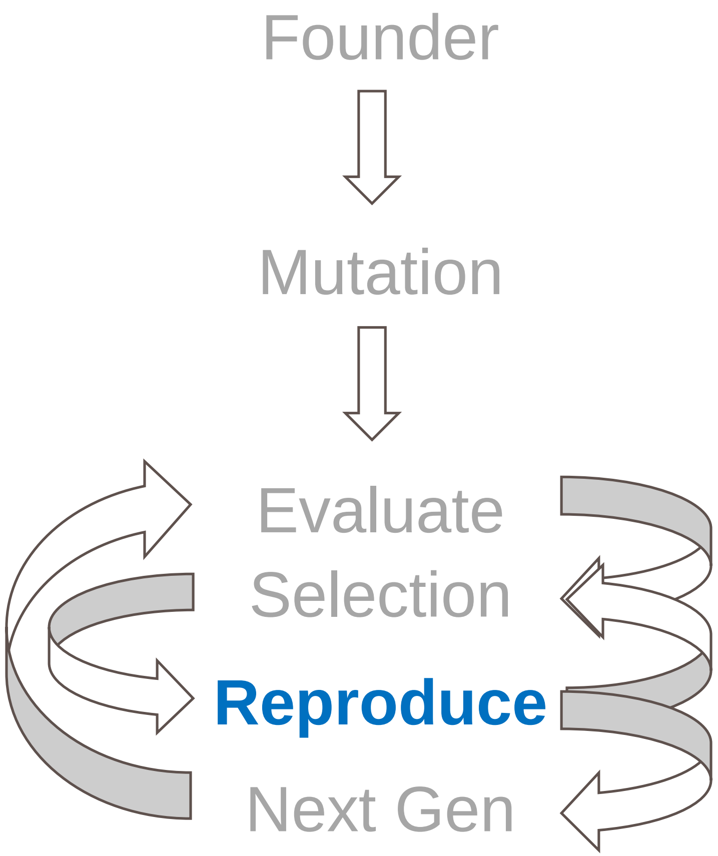
NN HPO – HPO with GAs: Mate Selection 1



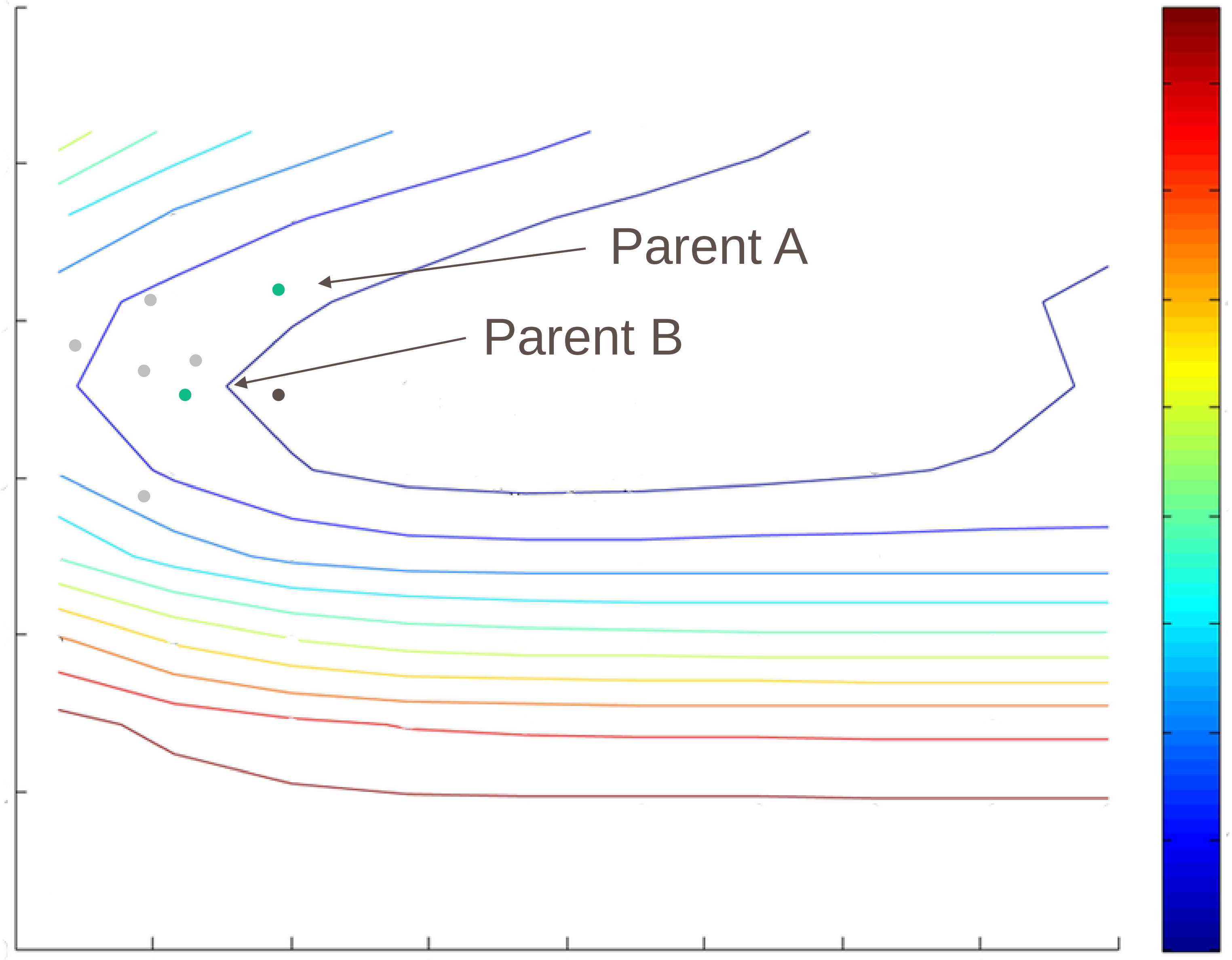


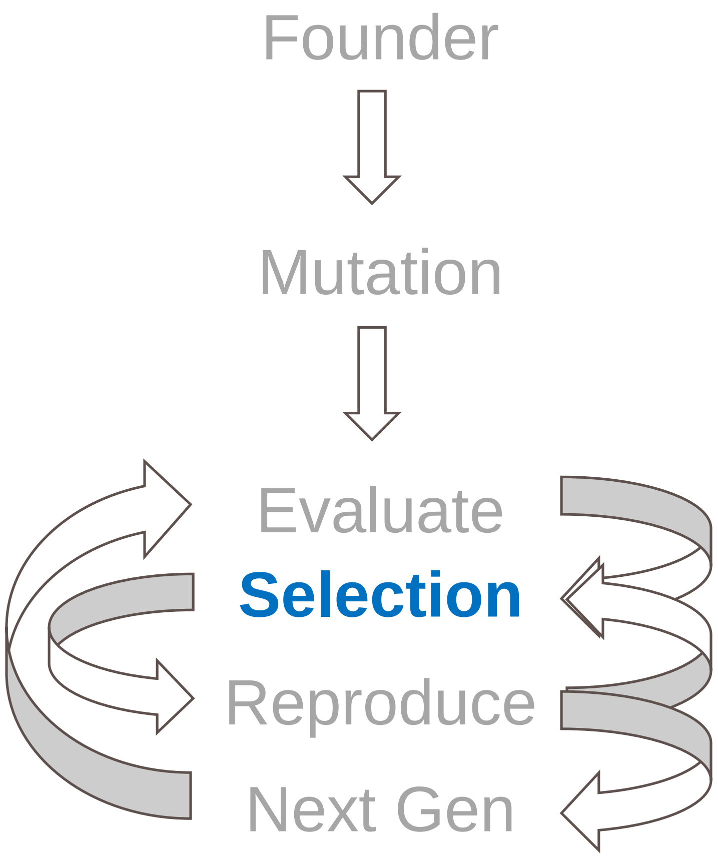
NN HPO – HPO with GAs: Reproduction 1



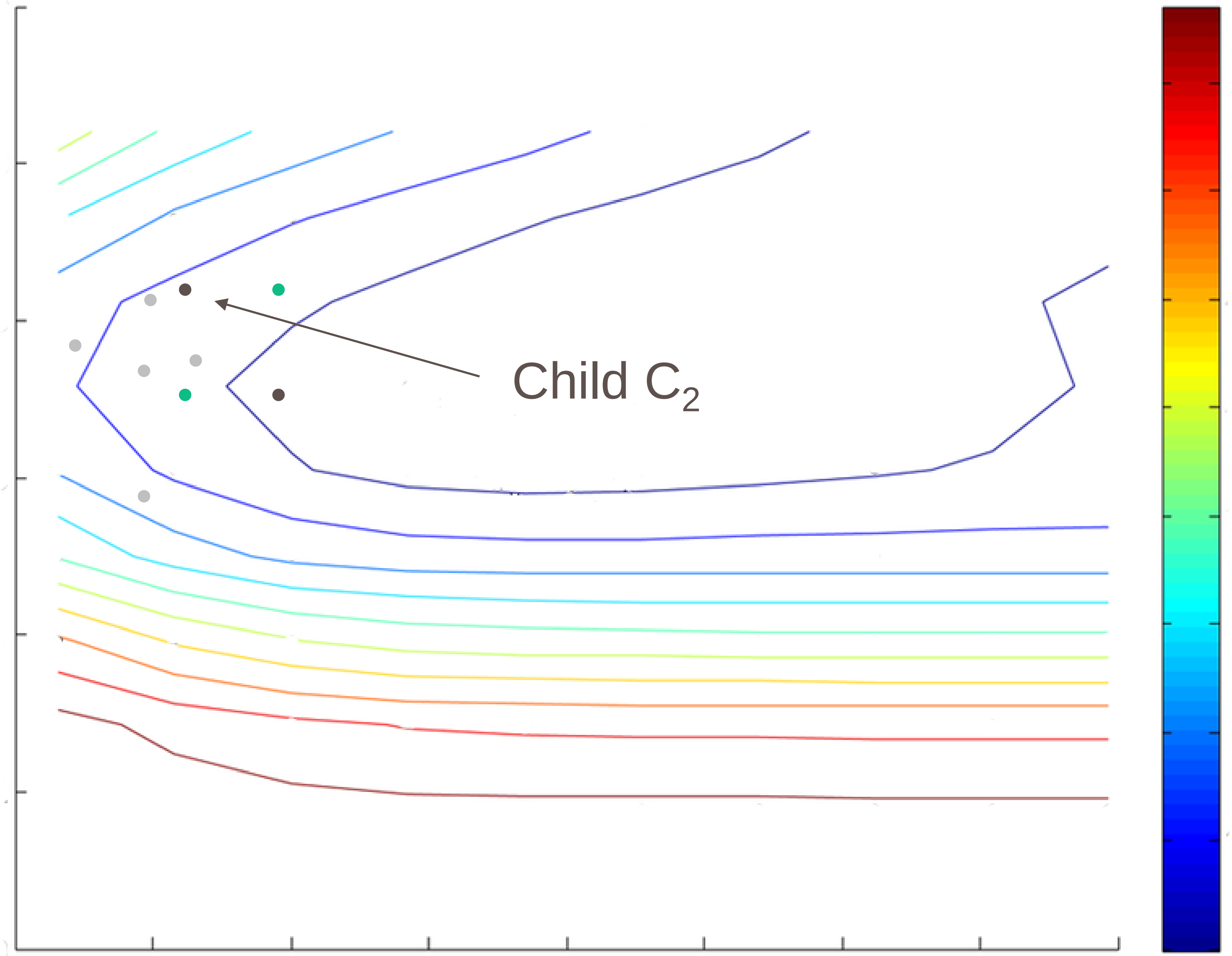


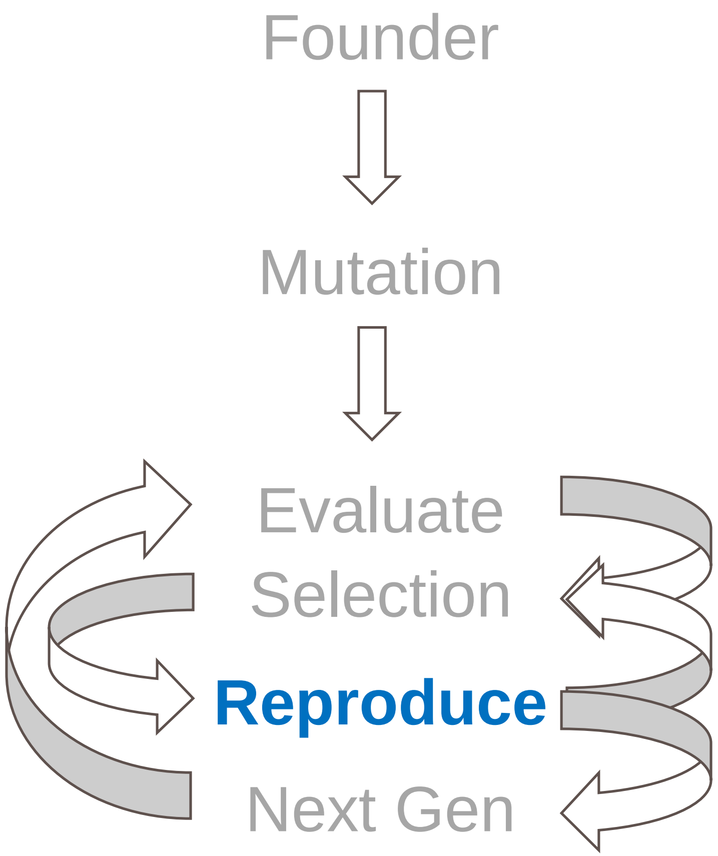
NN HPO – HPO with GAs: Mate Selection 2



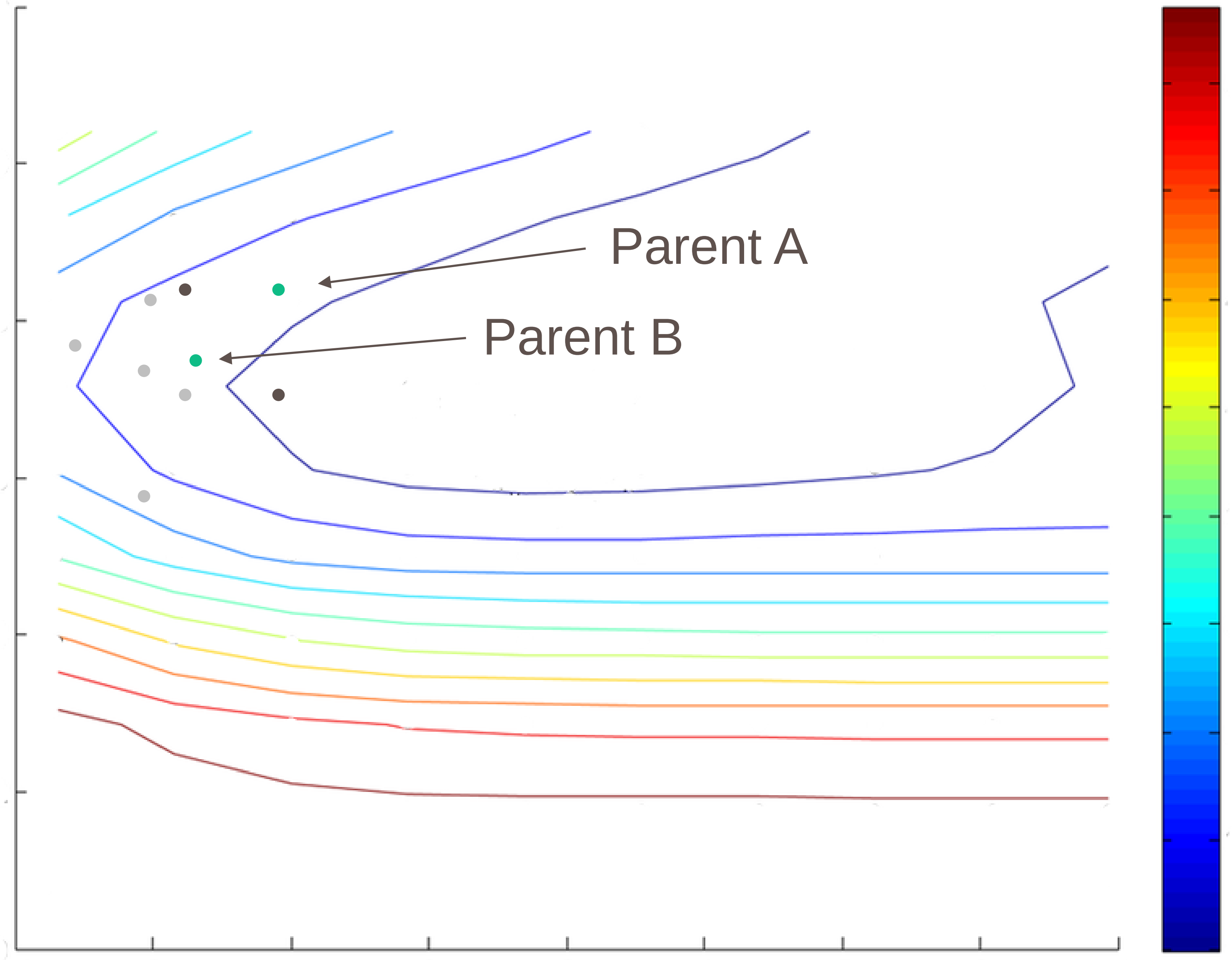


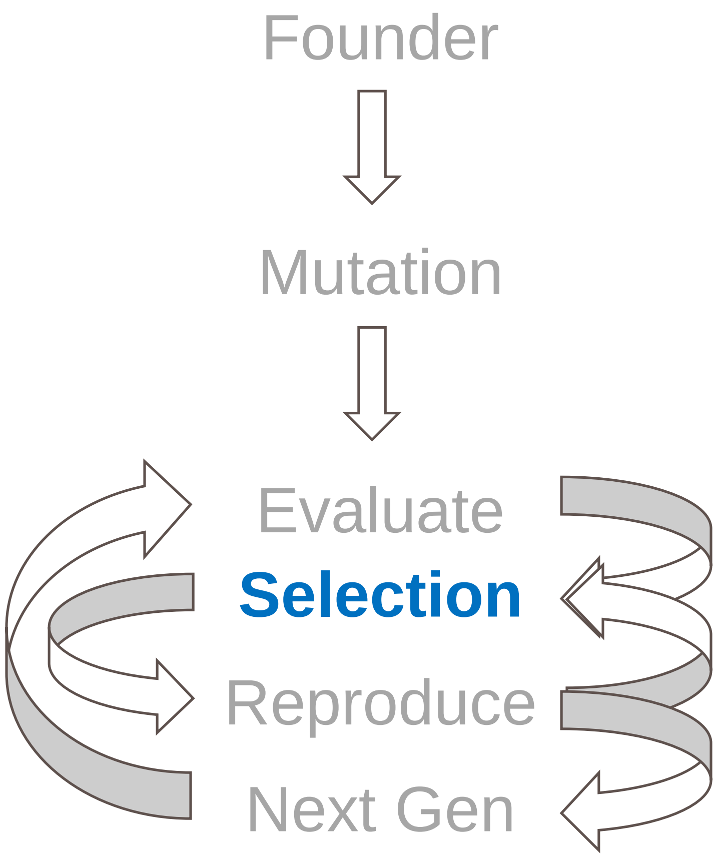
NN HPO – HPO with GAs: Reproduction 2



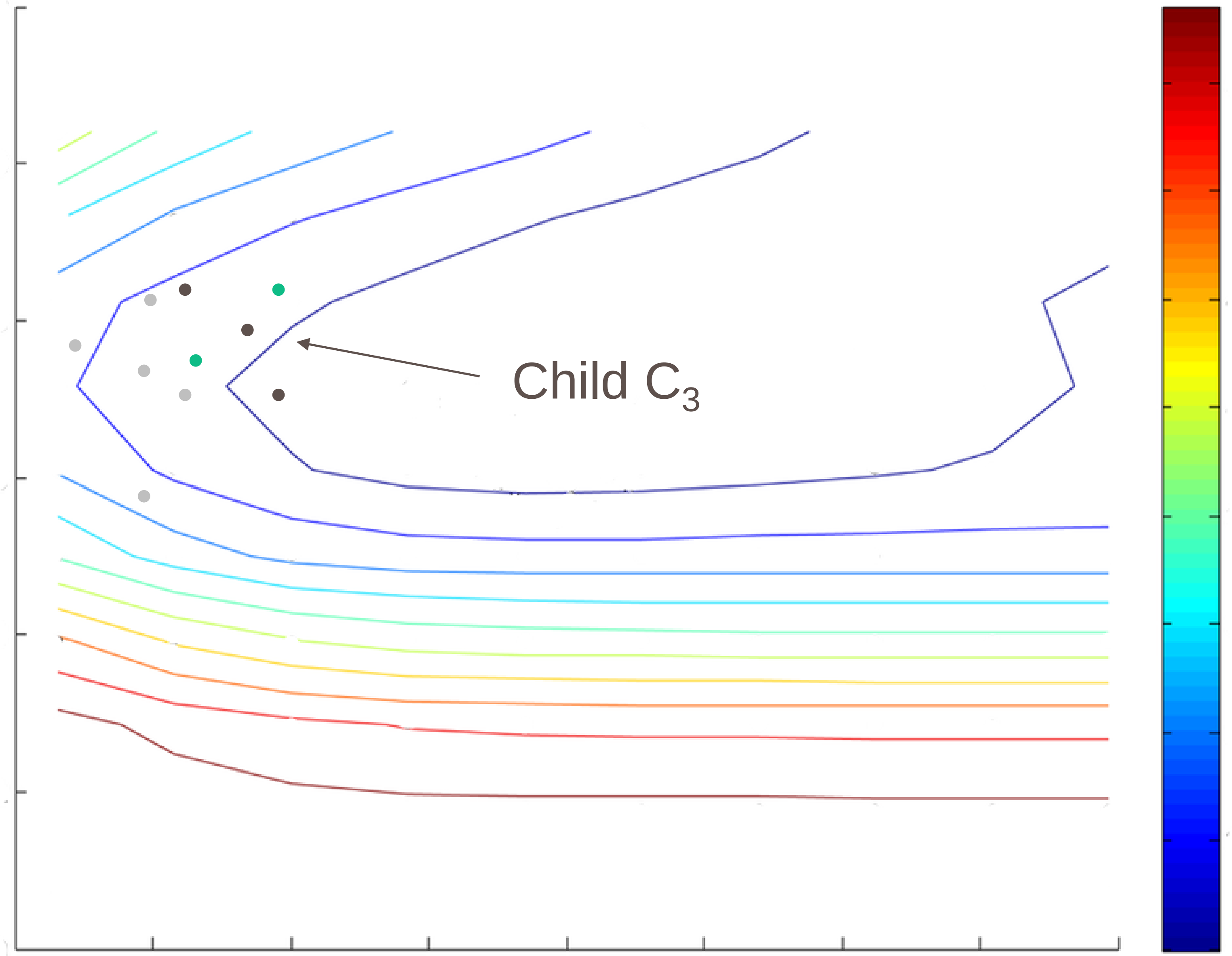


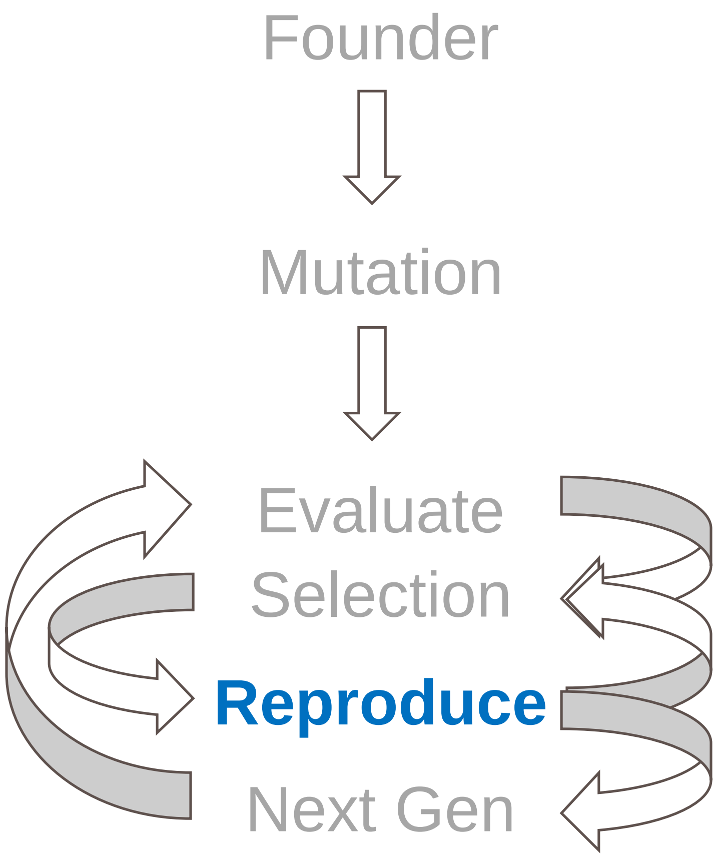
NN HPO – HPO with GAs: Mate Selection 3



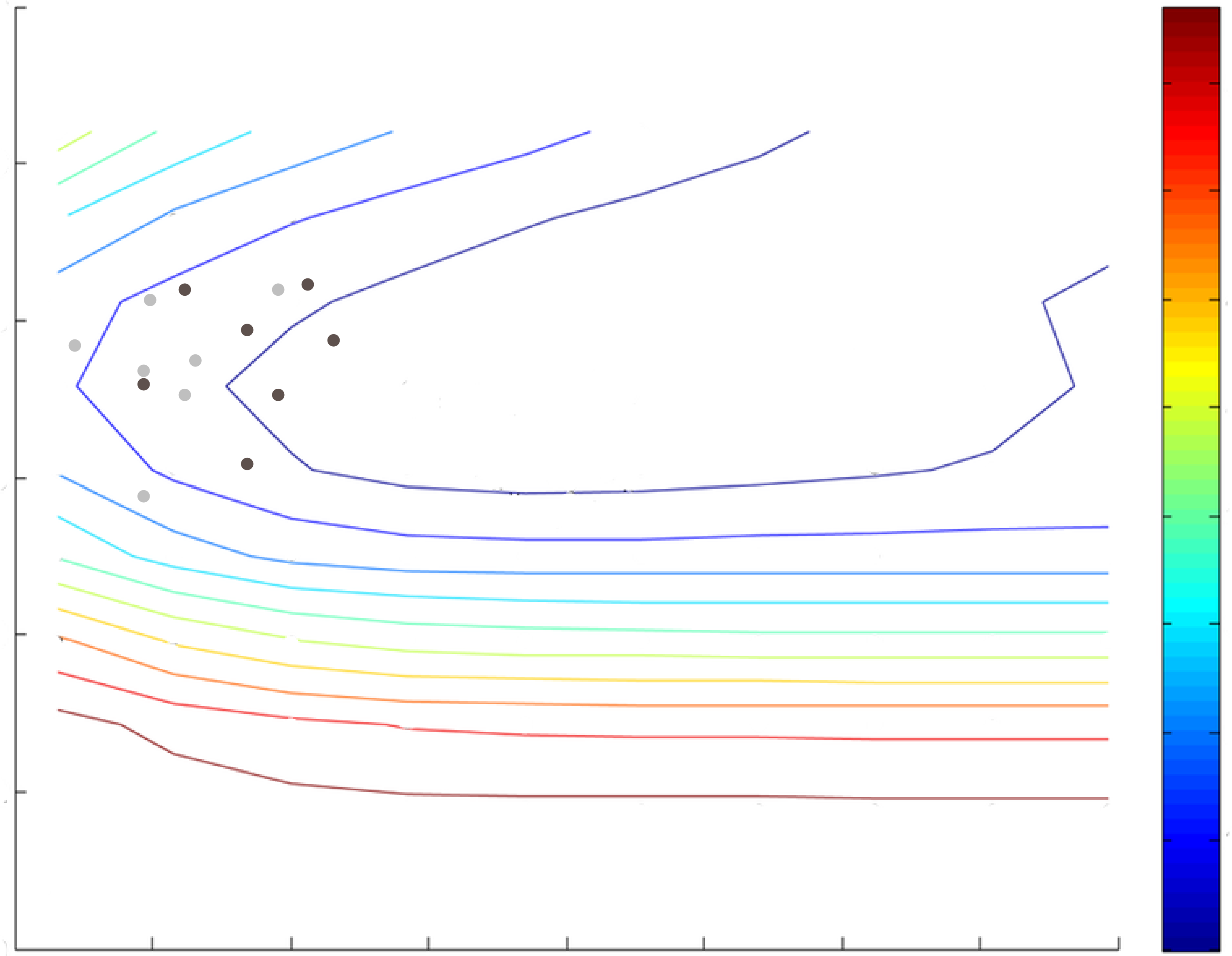


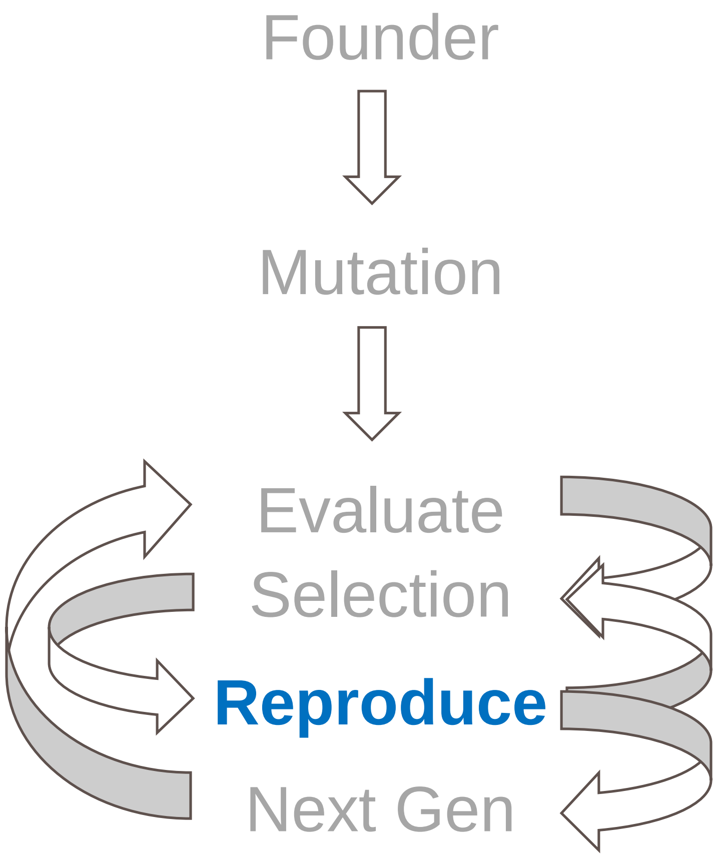
NN HPO – HPO with GAs: Reproduction 3



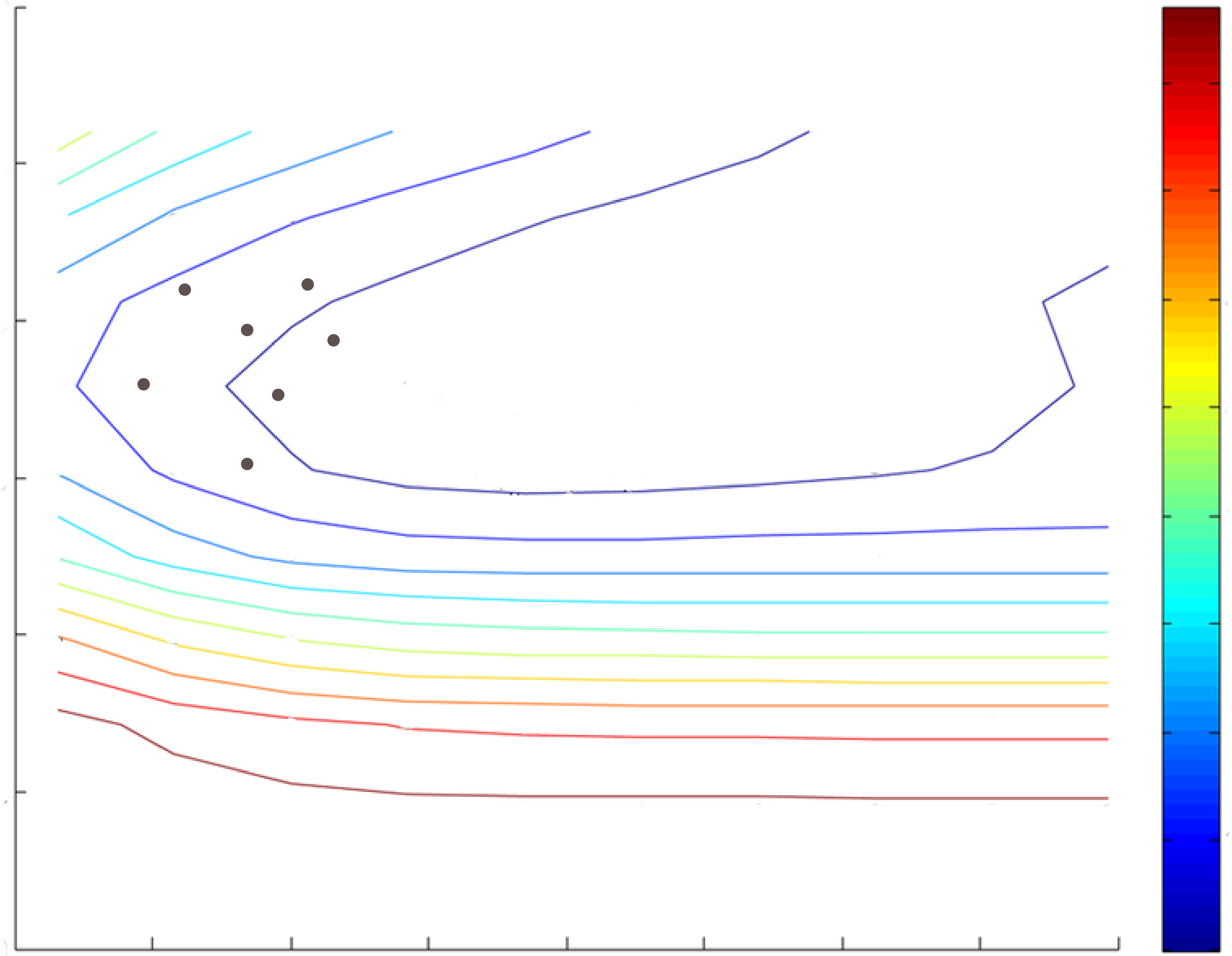


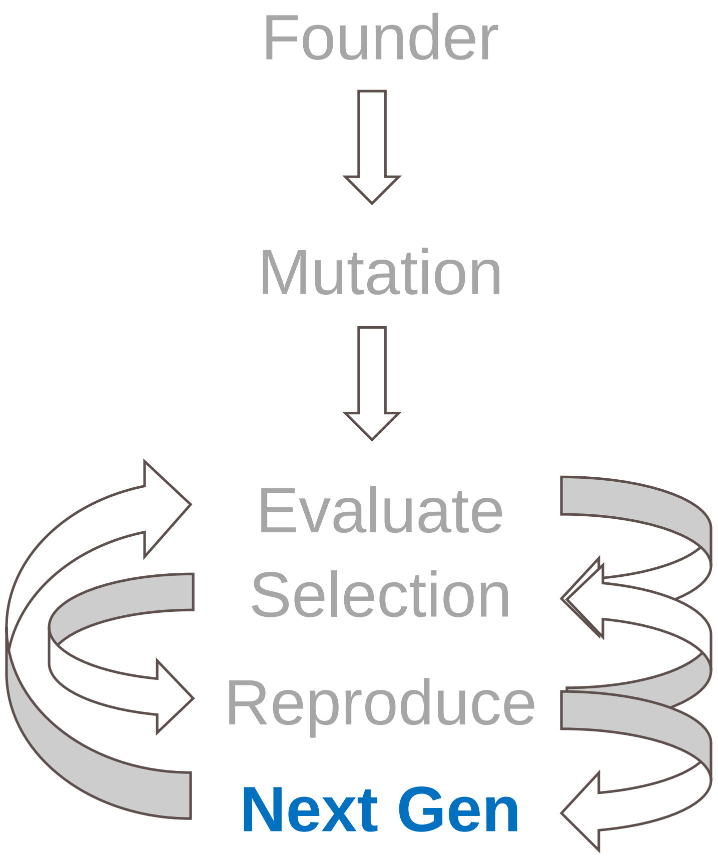
NN HPO – HPO with GAs: Reproduction N



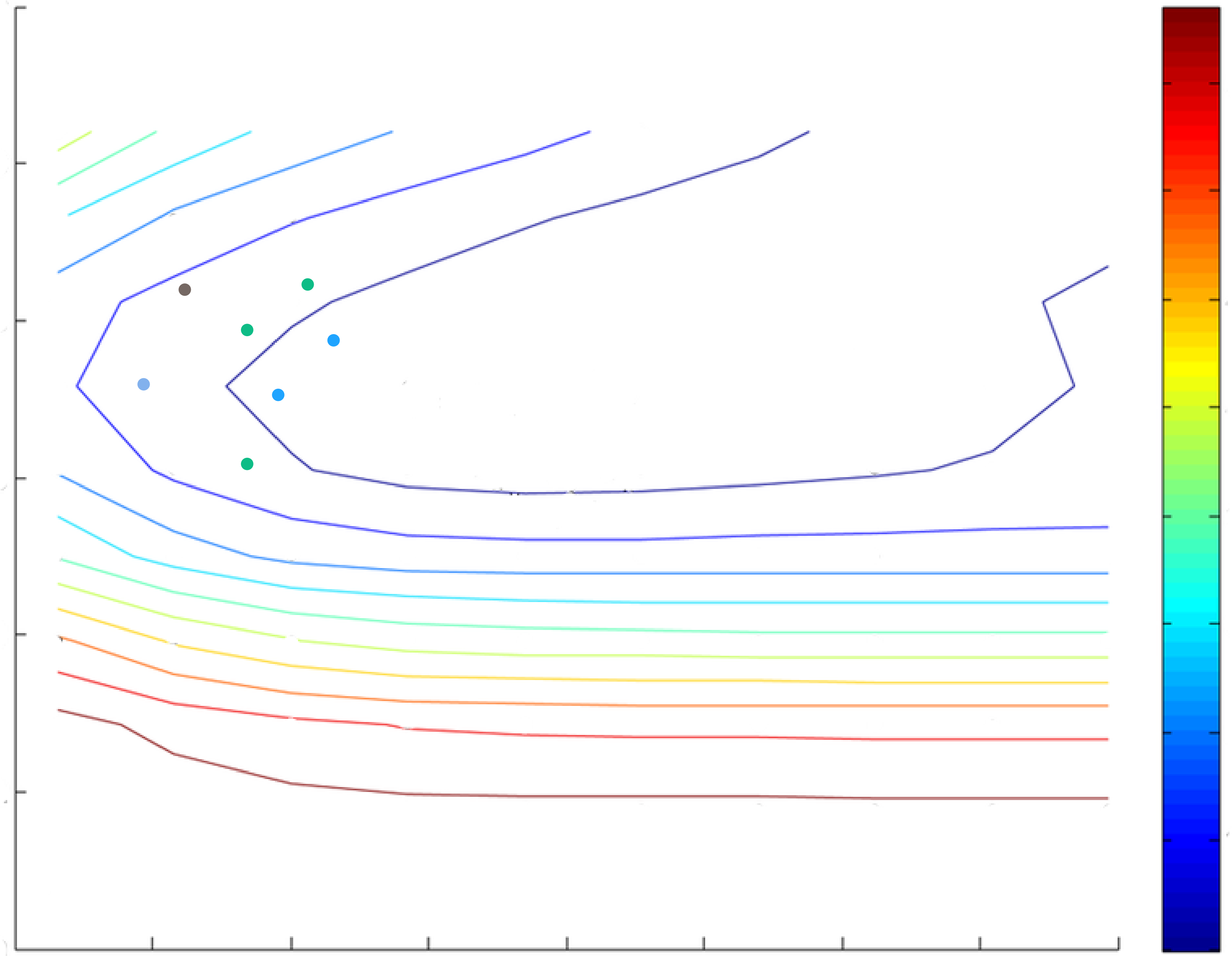


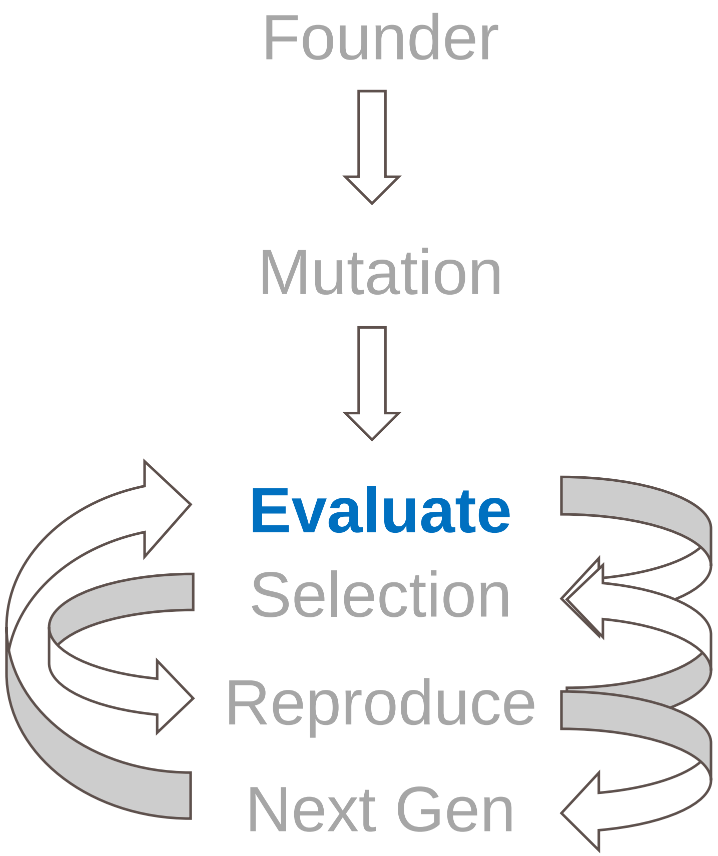
NN HPO – HPO with GAs: Next Generation



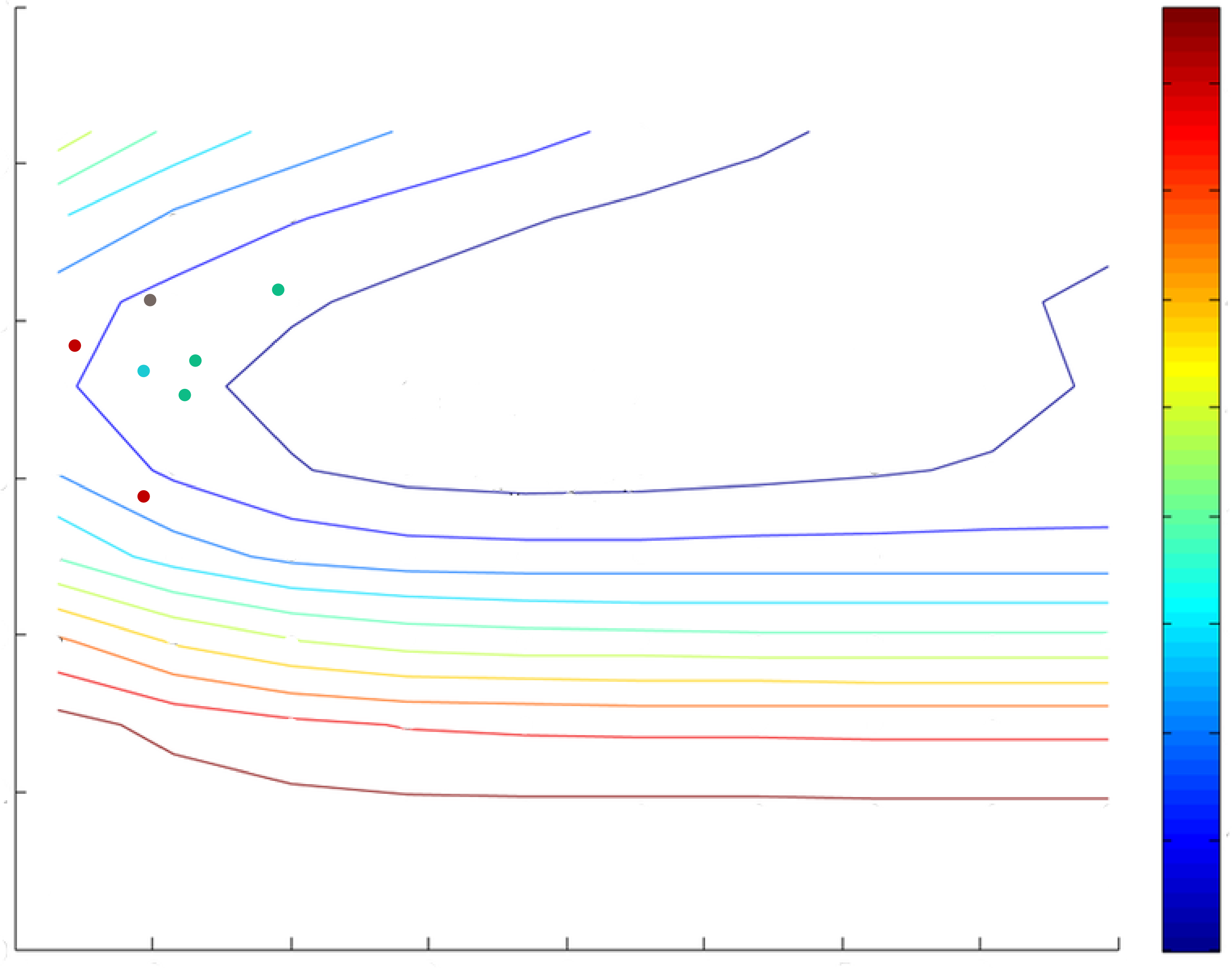


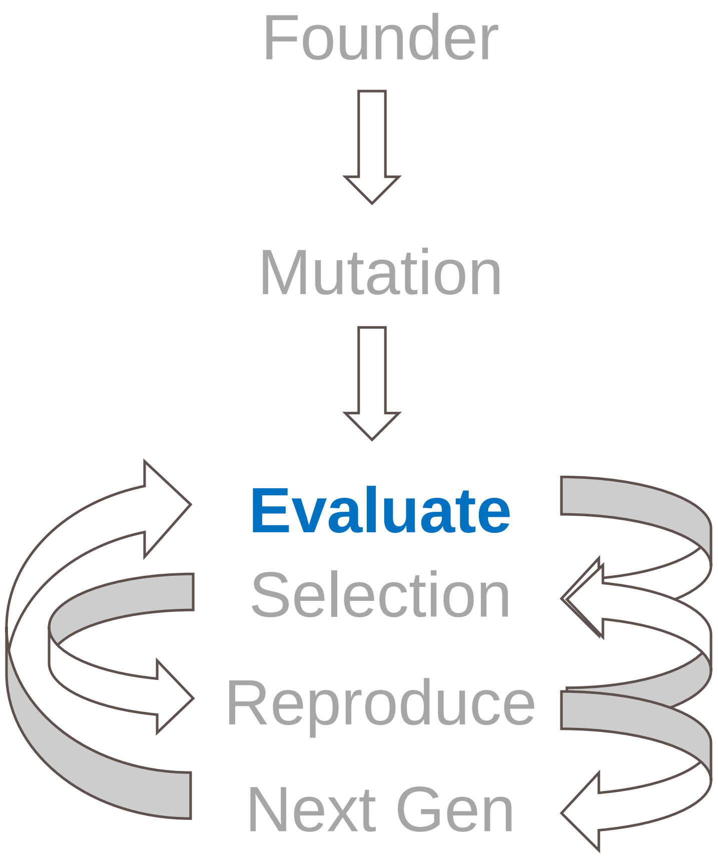
NN HPO – HPO with GAs: Evaluate Fitness 2



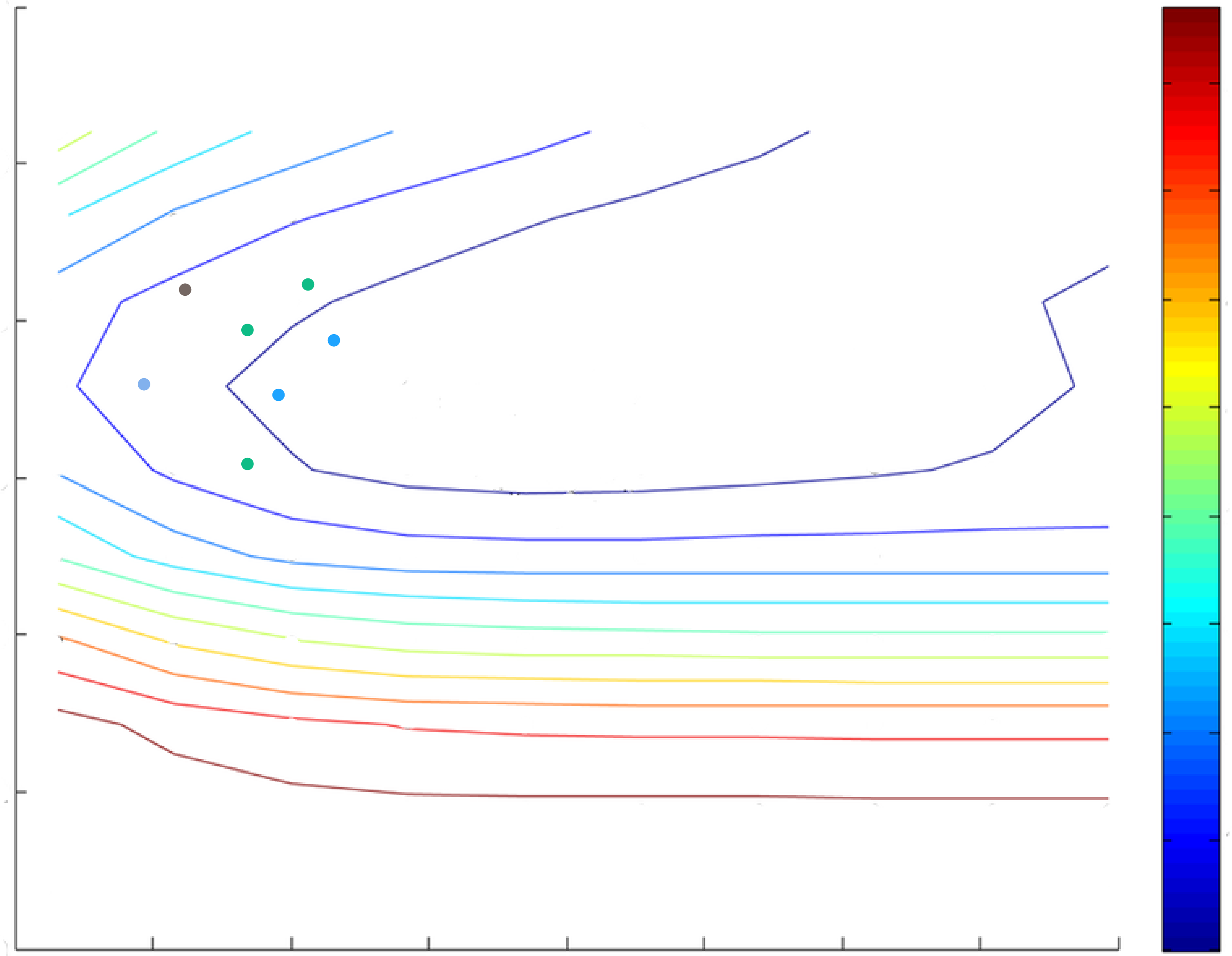


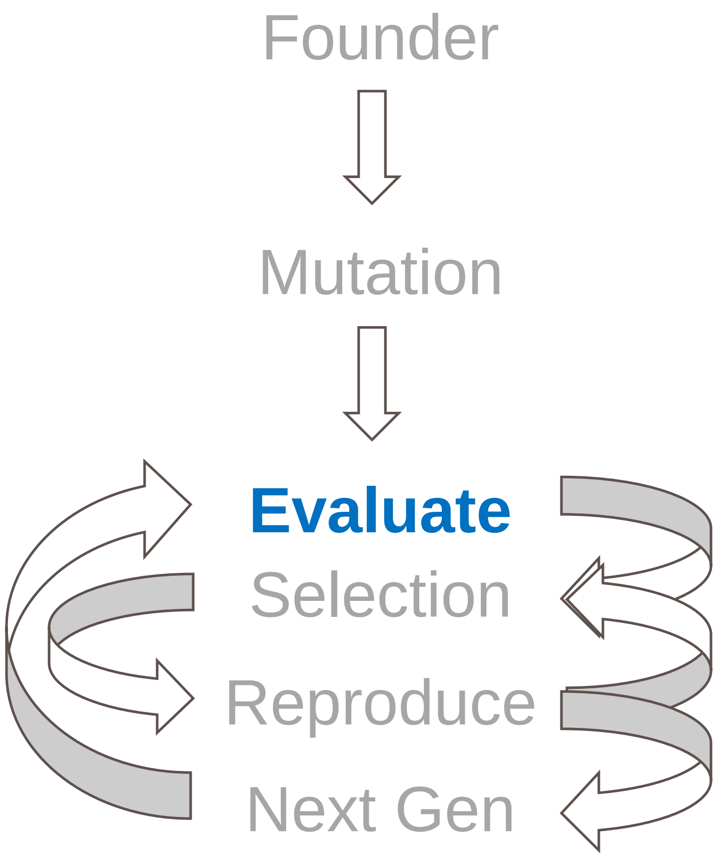
NN HPO – HPO with GAs: Evaluate Fitness 1



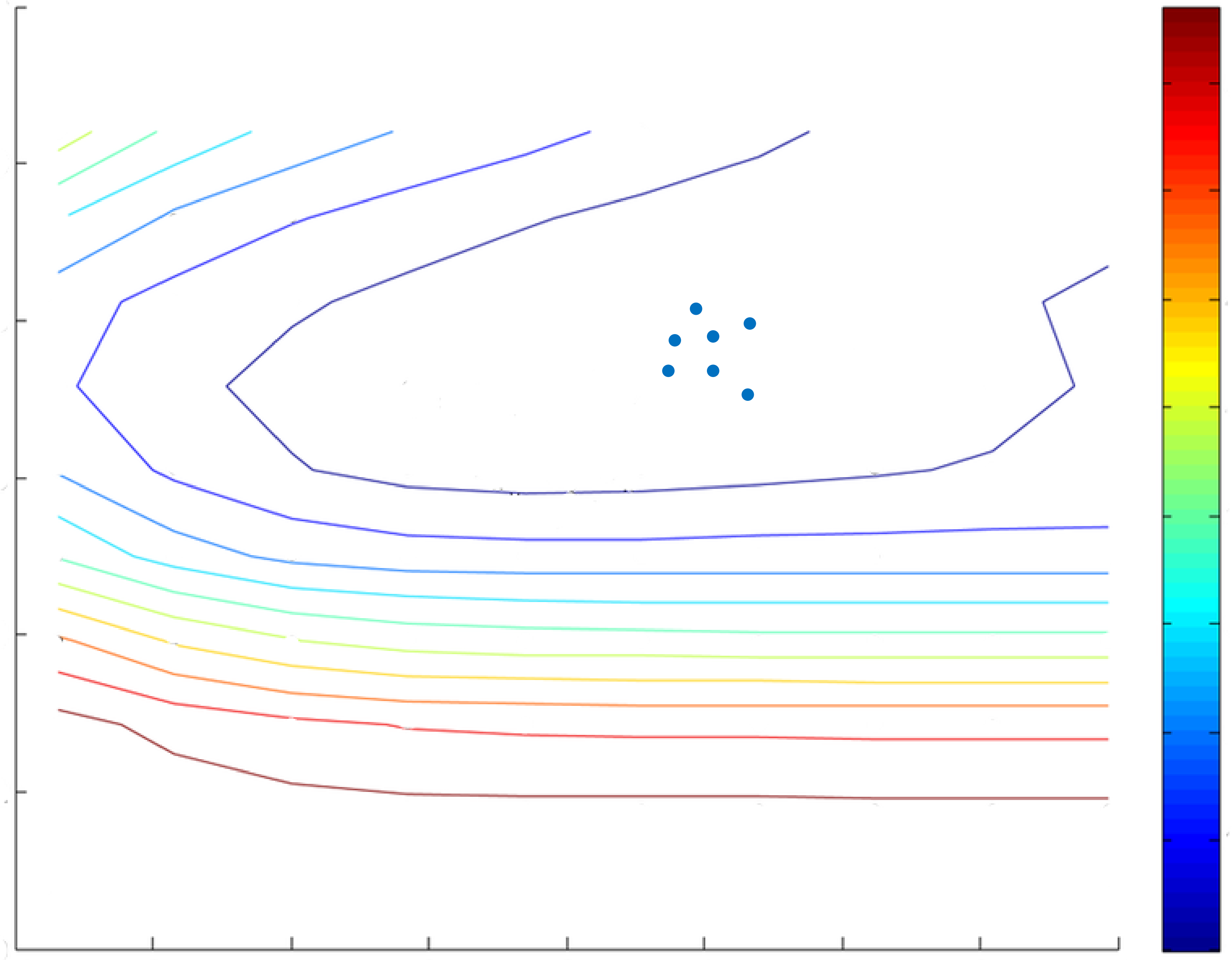


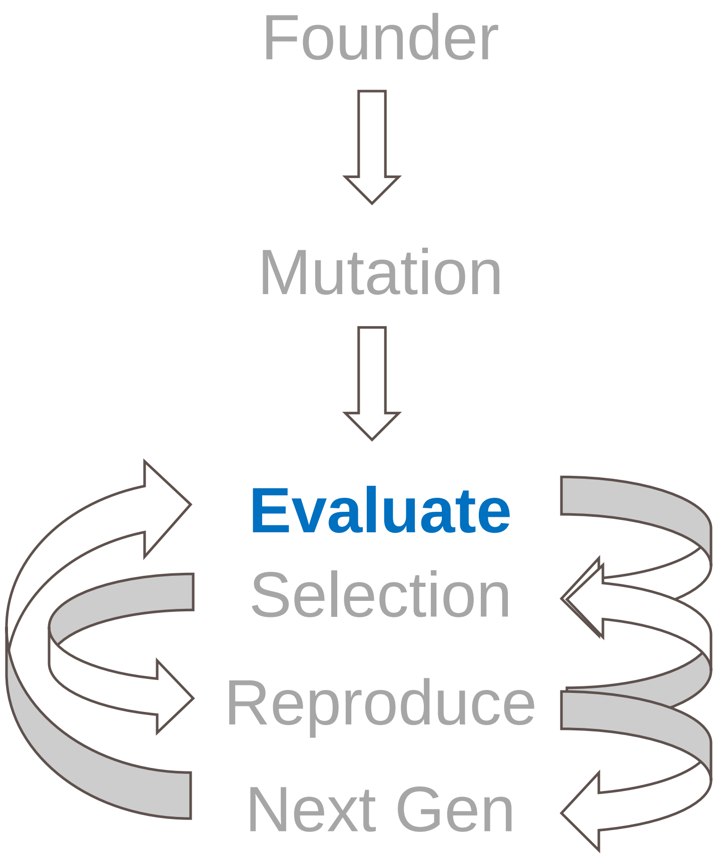
NN HPO – HPO with GAs: Evaluate Fitness 2





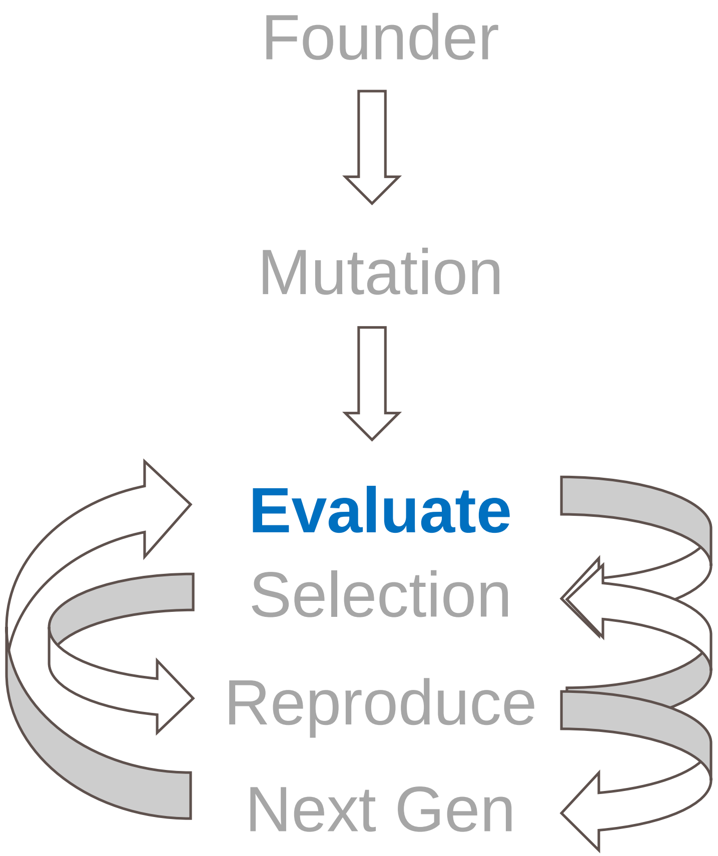
NN HPO – HPO with GAs: Evaluate Fitness N





NN HPO – HPO with GAs: Math

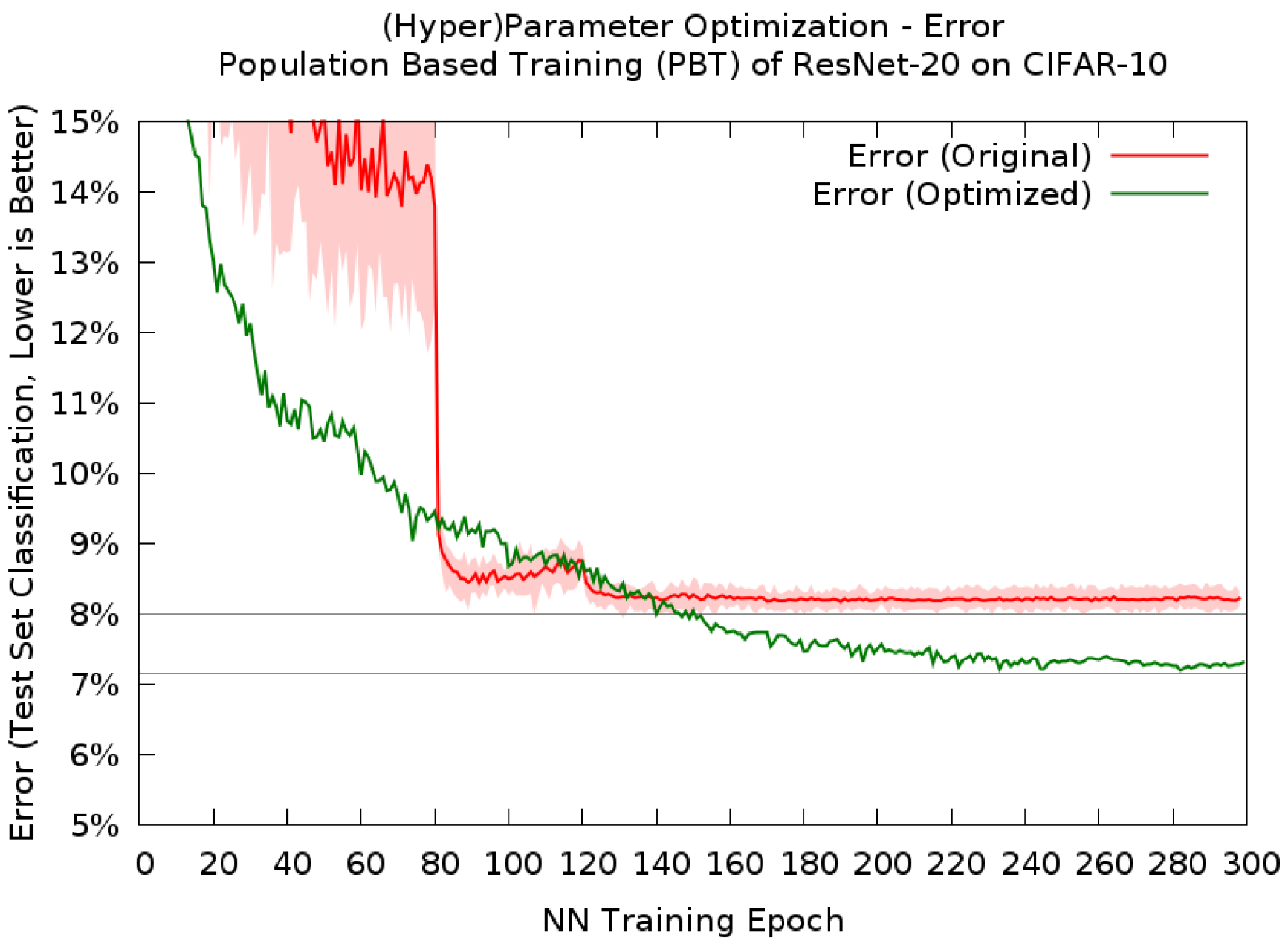
$g \leftarrow 0$
$\mathbb{P}_g \leftarrow$ initial_population
while $g <$ PARAM_GENERATIONS:
    for each p in $\mathbb{P}_g$:
        $p.fom \leftarrow$ execute(p)
        $p.fitness \leftarrow e^{-\sigma((p.fom - minfom)/(maxfom - minfom))^2}$
    $\mathbb{P}_{(g+1)} \leftarrow \emptyset$
    while $\left|\mathbb{P}_{(g+1)}\right| <$ PARAM_POPULATION_SIZE:
        $a \leftarrow$ choose p from $\mathbb{P}_g$ with probability proportional to $p.fitness$
        $b \leftarrow$ choose p from $\mathbb{P}_g$ with probability proportional to $p.fitness$
        $\alpha \leftarrow$ choose p from $\mathbb{P}_g$ with probability proportional to $p.fitness$
        $c.hparams \leftarrow$ mutate(crossover($a.hparams,\ b.hparams$))
        $c.params \leftarrow \alpha.params$
        $\mathbb{P}_{(g+1)} \leftarrow \mathbb{P}_{(g+1)} \cup \{c\}$



Put It All Together

NN HPO – Population-Based Training

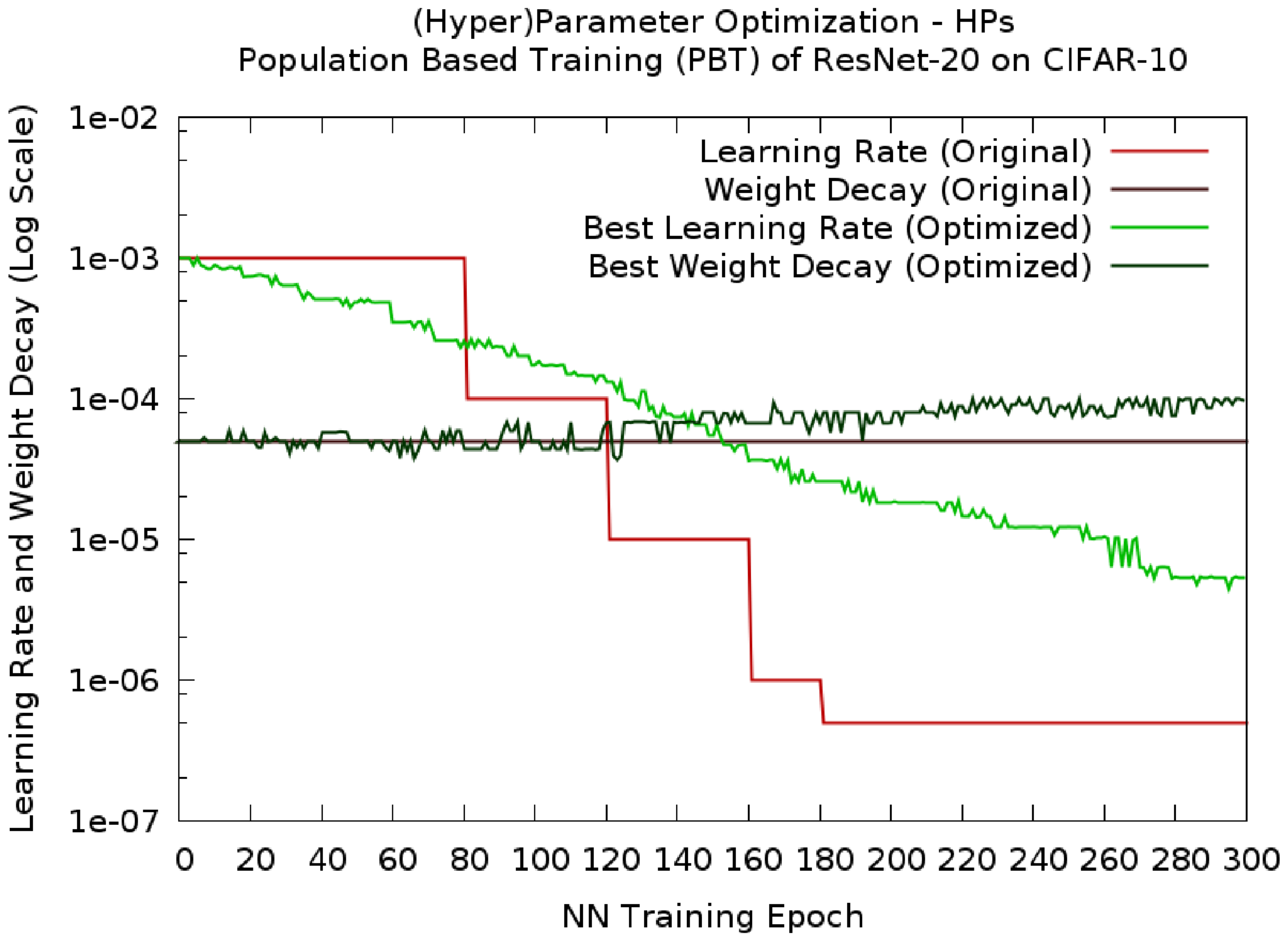
- Population-based training (PBT): interleave SGD and GA-based HPO:
  - Train all NNs in population for one epoch.
  - Save NN model weights and note accuracy for fitness.
  - Treat NN weights as a gene in the GA during reproduction.



https://arxiv.org/pdf/1901.03900.pdf
https://arxiv.org/pdf/1711.09846.pdf

NN HPO – PBT Provides HP Training Schedule

- Population-based training (PBT): interleave SGD and GA-based HPO:
  - Train all NNs in population for one epoch.
  - Save NN model weights and note accuracy for fitness.
  - Treat NN weights as a gene in the GA during reproduction.



https://arxiv.org/pdf/1901.03900.pdf
https://arxiv.org/pdf/1711.09846.pdf

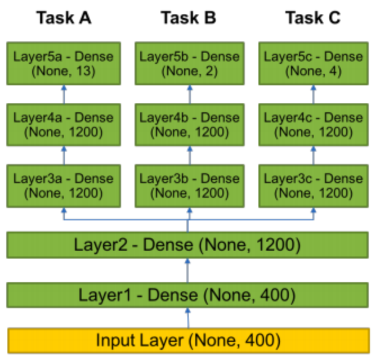
NN HPO – Population-Based Training Results

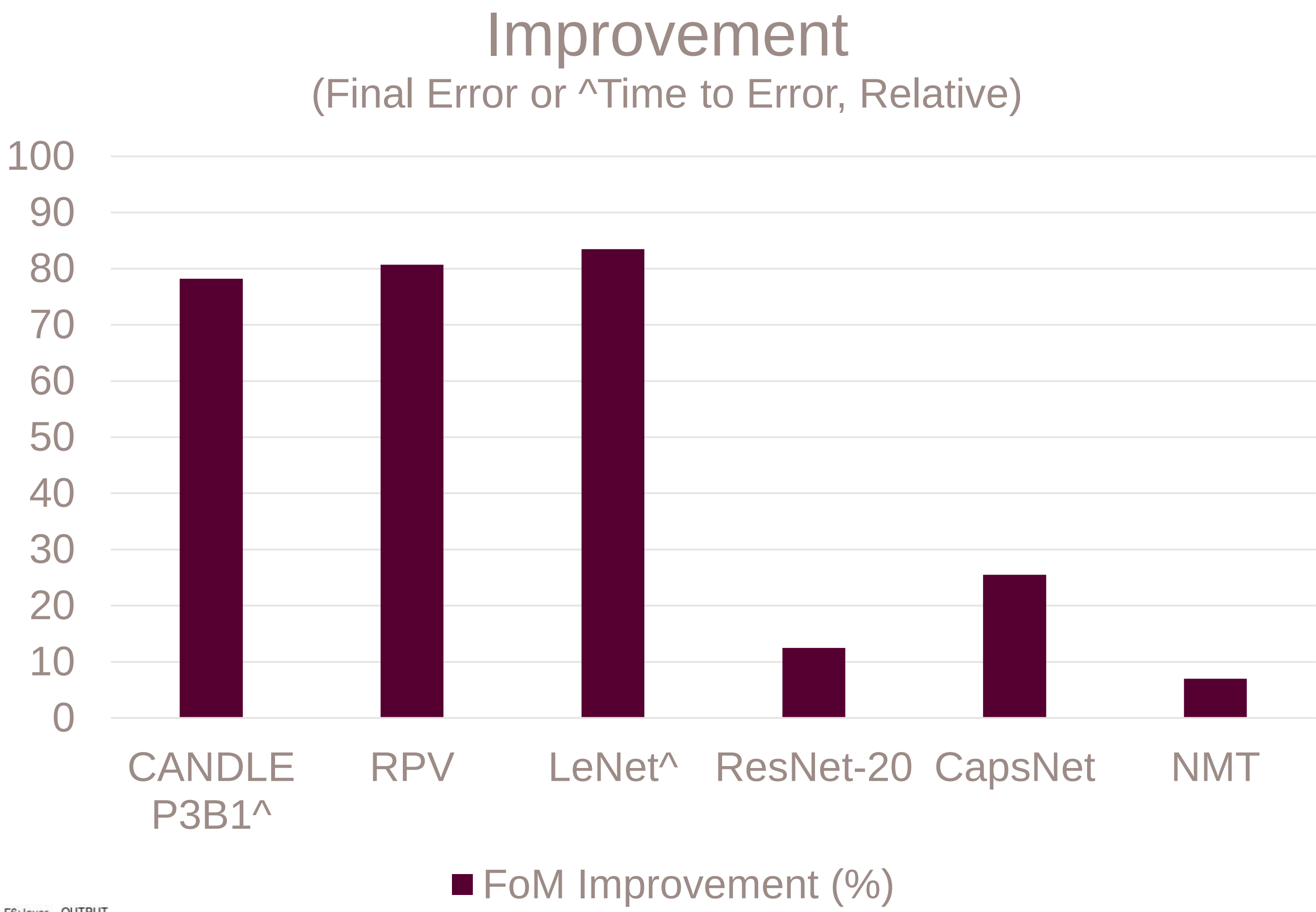
- Population-based training (PBT): interleave SGD and GA-based HPO:
  - Train all NNs in population for one epoch.
  - Save NN model weights and note accuracy for fitness.
  - Treat NN weights as a gene in the GA during reproduction.

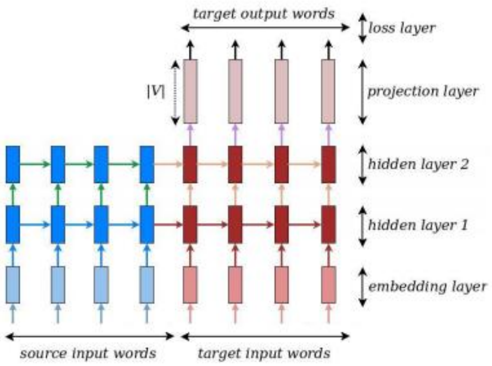
Model w/ CIFAR10	Size (MParam)	Implementation Accuracy	Original Paper^
ResNet-20 (HPO:PBT)	**0.27**	**93.00%**	--
ResNet-20	0.27	92.16%	91.25%
ResNet-32	0.46	92.46%	92.49%
ResNet-44	0.66	92.50%	92.83%
ResNet-56	0.85	92.71%	93.03%
ResNet-110	1.70	92.65%	93.39%

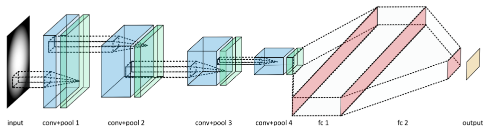
https://arxiv.org/pdf/1901.03900.pdf
^https://arxiv.org/pdf/1512.03385.pdf

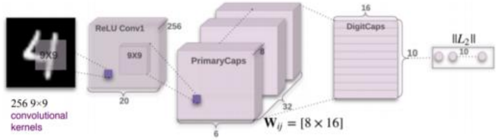
NN HPO – HPO Results

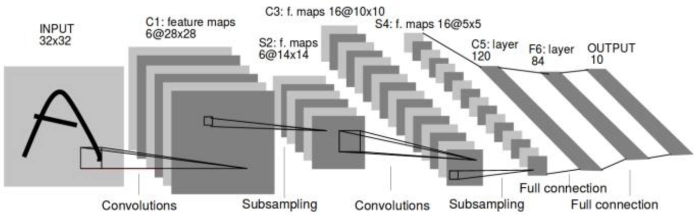


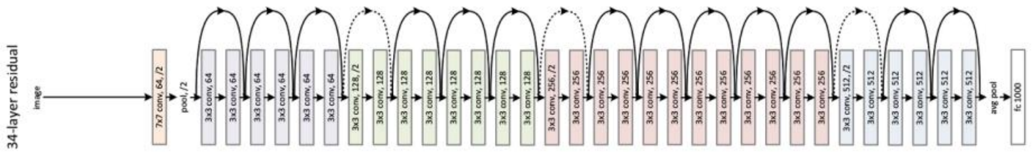












https://arxiv.org/pdf/1901.03900.pdf

QUESTIONS?

avose@cray.com

SAFE HARBOR STATEMENT

This presentation may contain forward-looking statements that are based on our current expectations. Forward looking statements may include statements about our financial guidance and expected operating results, our opportunities and future potential, our product development and new product introduction plans, our ability to expand and penetrate our addressable markets and other statements that are not historical facts.

These statements are only predictions and actual results may materially vary from those projected. Please refer to Cray's documents filed with the SEC from time to time concerning factors that could affect the Company and these forward-looking statements.