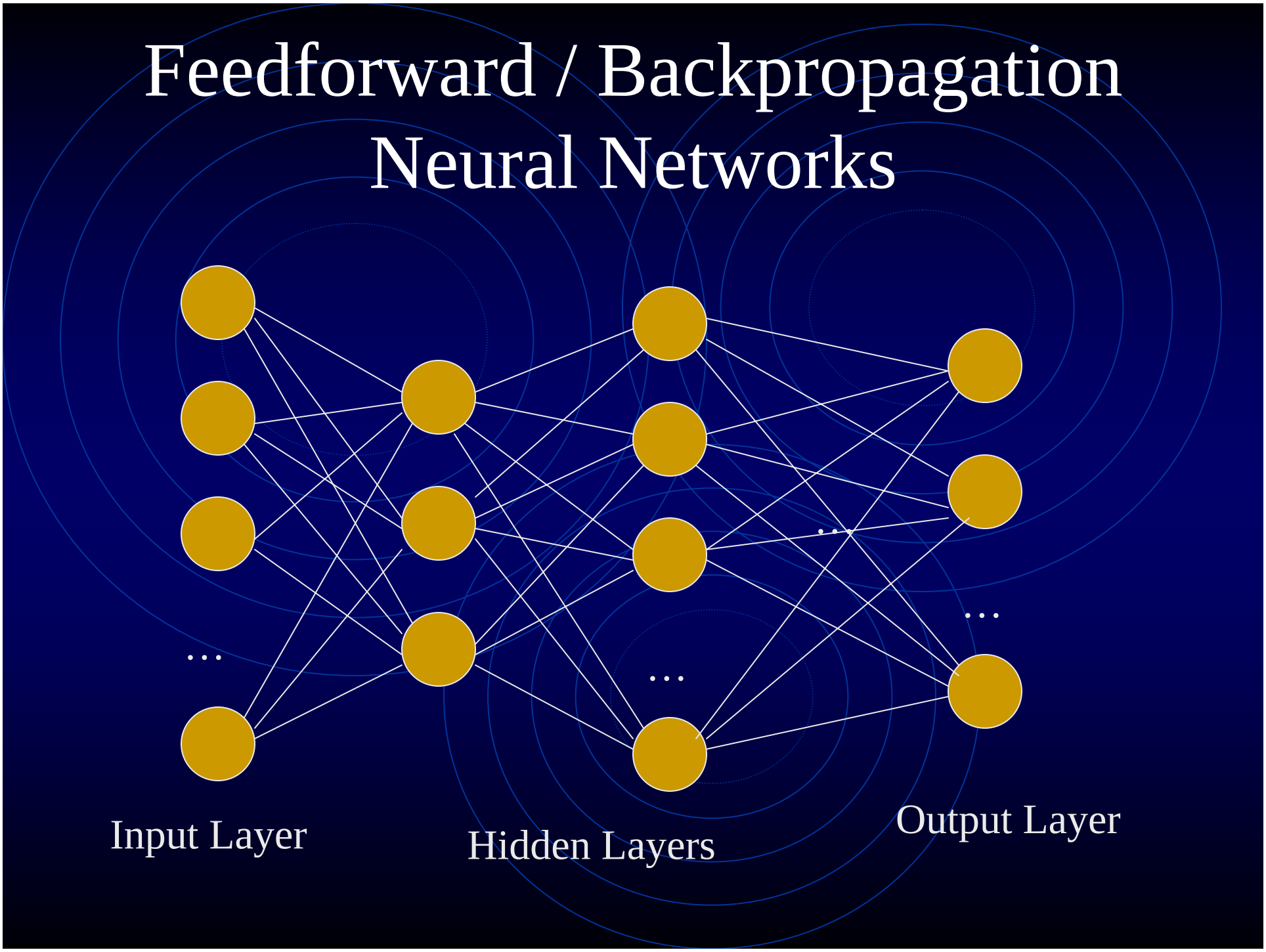
Comparing Standard and Flat Cascade-Correlation Neural Networks

Frédéric Dandurand
Chulalongkorn University
Bangkok, Thailand
January 2005

Neural Networks in Cognitive Psychology

- Models of cognitive processes
  - Visual perception
  - Associative memory
- Features
  - Ability to learn
  - Processing of degraded (e.g. noisy) signals
  - Fault tolerant
  - Biologically inspired and plausible

Feedforward / Backpropagation Neural Networks



Feedforward / Backpropagation Neural Networks

- Problem: Set network architecture parameters
  - Number of hidden layers
  - Number of hidden units
- Based on knowledge of the problem and heuristics → Not cognitively plausible

Constructive Neural Networks

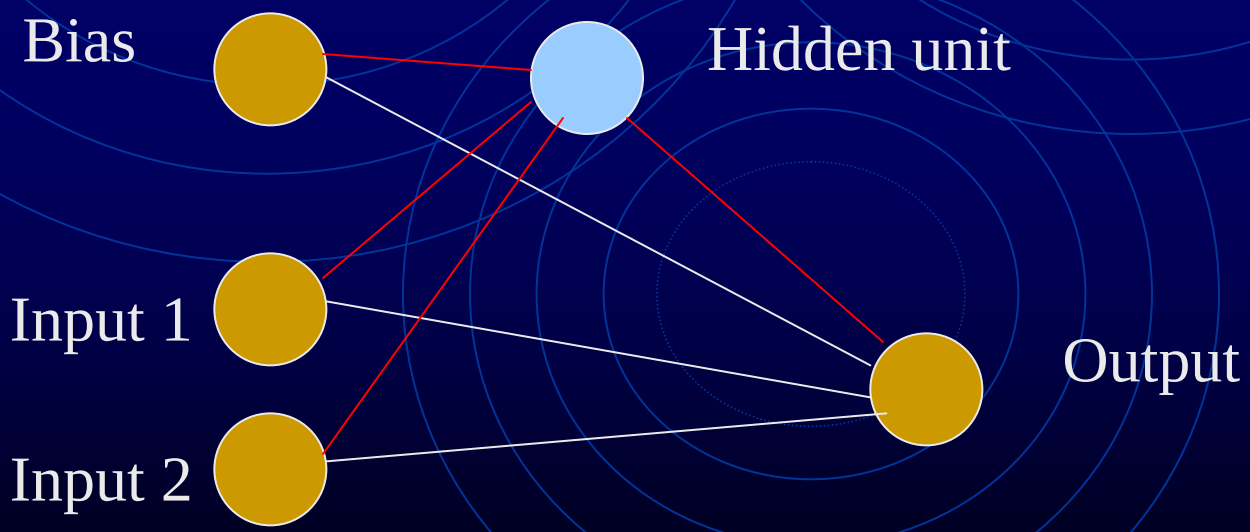
- Family of network architectures that grow as they learn
- Solution to the dimensioning problem
- Idea: start small and simple, and add processing power (unit and/or layer) if necessary
- Cascade-Correlation (Fahlman 1990)

Cascade-Correlation

- Example: 2 inputs, 1 output
- Initial architecture

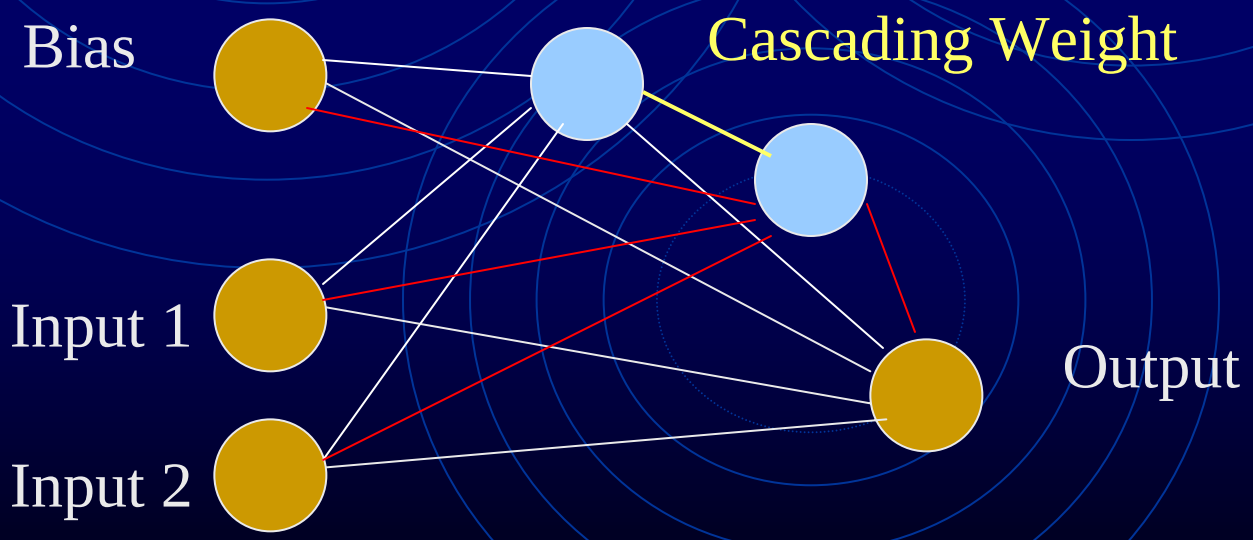
Cascade-Correlation

- Hidden unit recruitment



Cascade-Correlation

- Second hidden unit recruitment

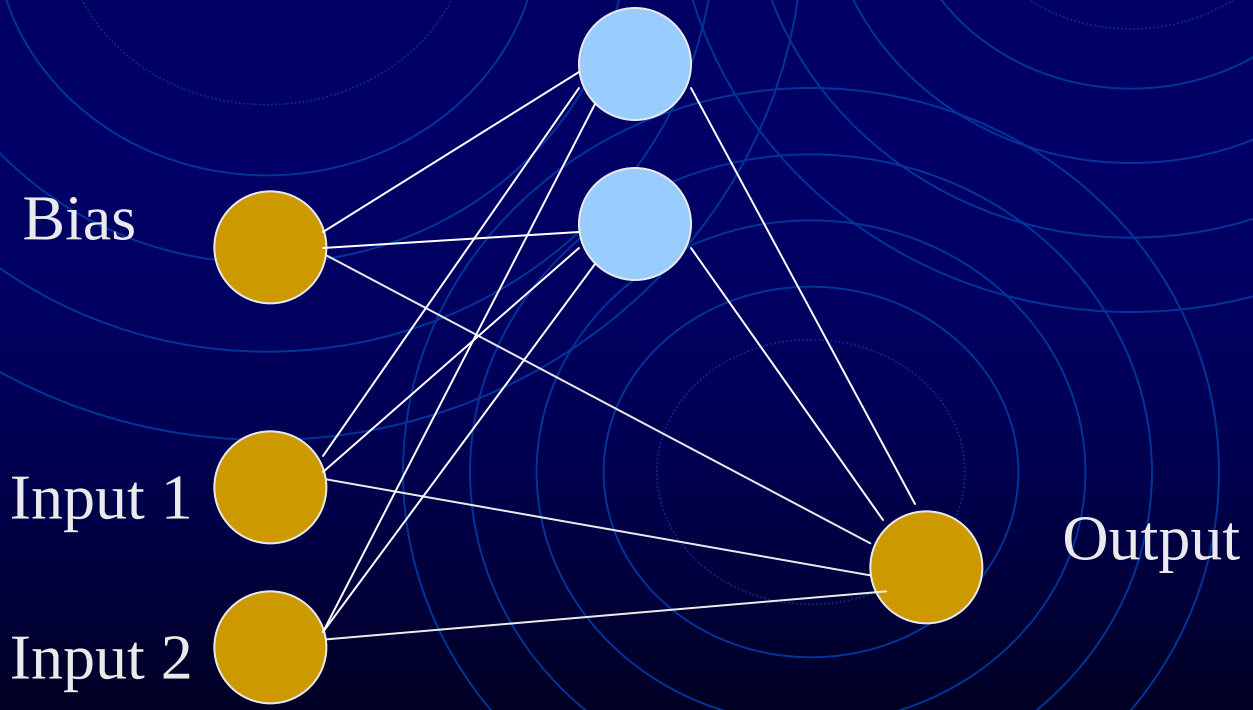


Cascading Weights

- Standard Cascade-Correlation (Fahlman) has cascading weights
- Layers count = recruitment count
- Effect: create « deep » networks
- Critics: bias toward non-linearity → Harms generalization ability

Flat Cascade-Correlation

- Eliminate cascading weights
- « Flat » Cascade-correlation networks

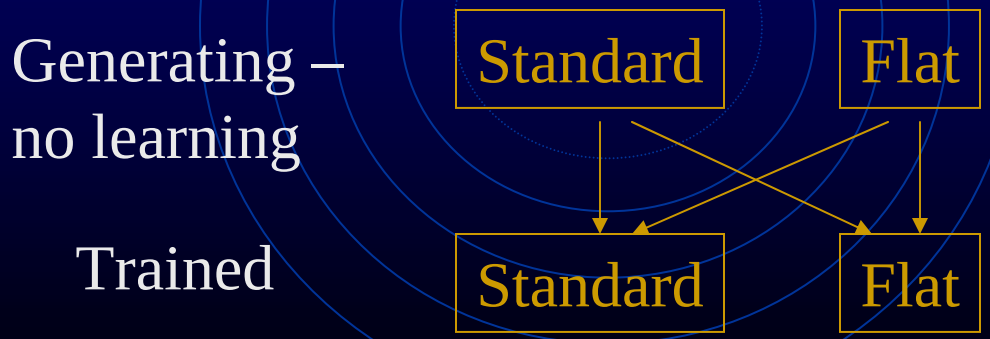


Standard Vs. Flat Cascade-Correlation

- Which one is better? Standard (Deep) or Flat
- No consensus in the litterature
  - Prechelt (1997)
    - Extensive testing using real-life problems (PROBEN)
    - Equivalent performance of standard and flat most of the time
    - When not, flat was better more often than the reverse
  - Littmann & Ritter (1993)
    - Standard networks generalize better than flat networks when the training set is small

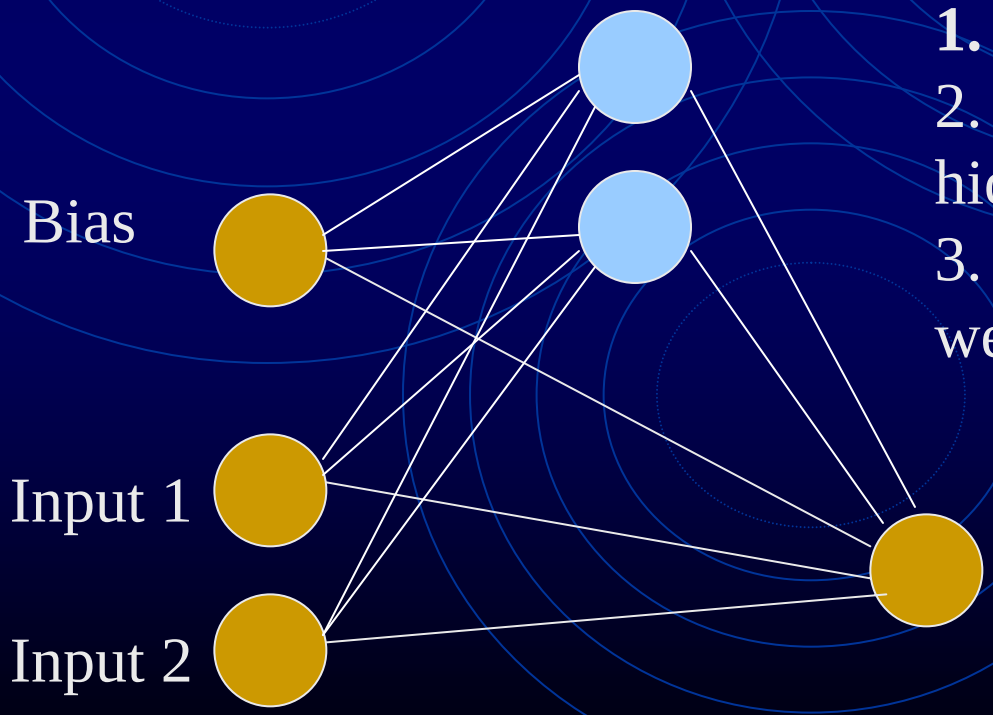
Network Output Approximation Experiment

- Instead of using a problem set, directly compare the ability of standard networks to approximate flat networks and vice-versa
- Generate data using standard and flat networks
- Train standard and flat networks



Experimental Setup

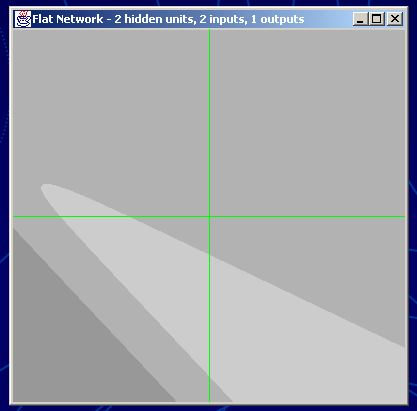
- 2 inputs, 1 output
- Vary the number of hidden units: 2 to 50

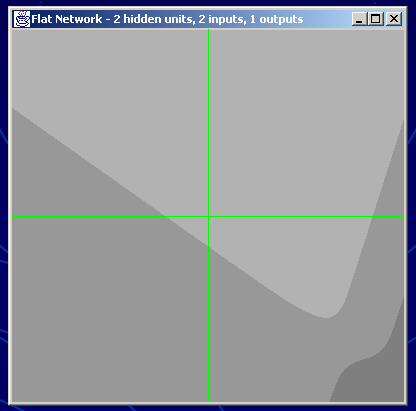


1. **No learning**
2. Manually install hidden units
3. Random connection weights

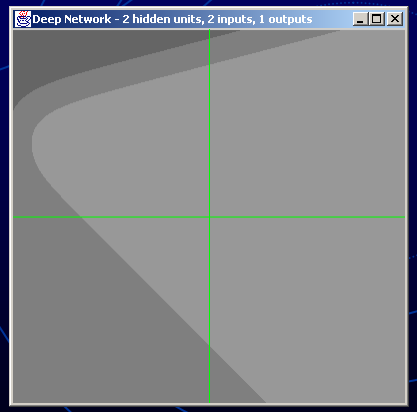
Network Output – 2 hidden units

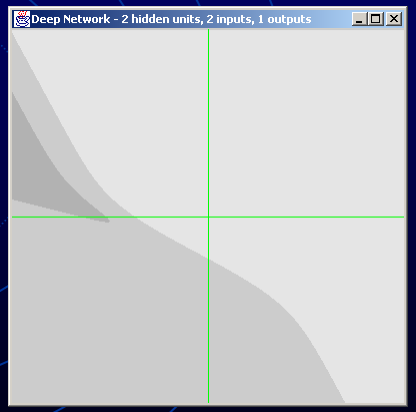
Flat





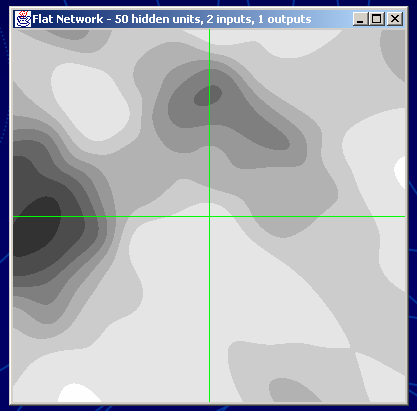
Standard

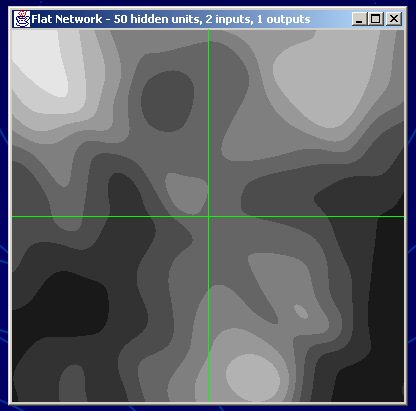




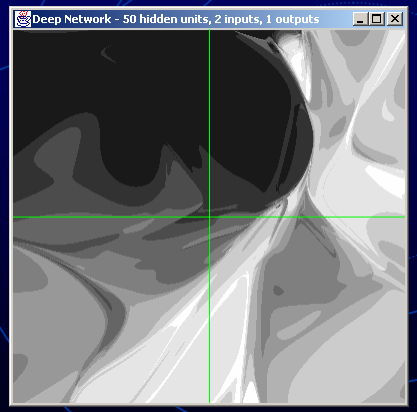
Network Output – 50 hidden units

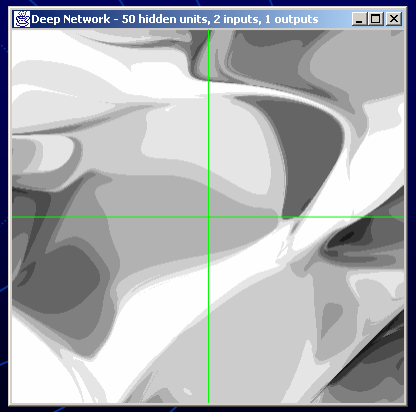
Flat





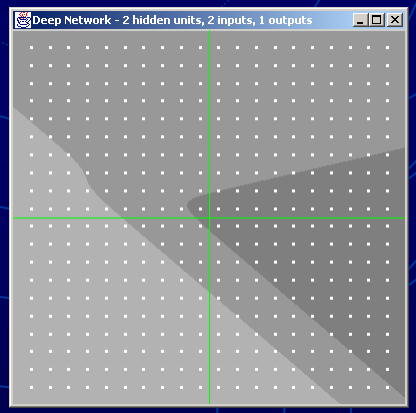
Standard -
Bias toward
Non-linearity

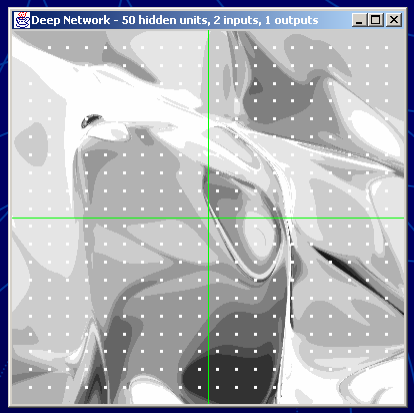




Network Output Sampling

- To generate the training set, sample the output function over 400 points





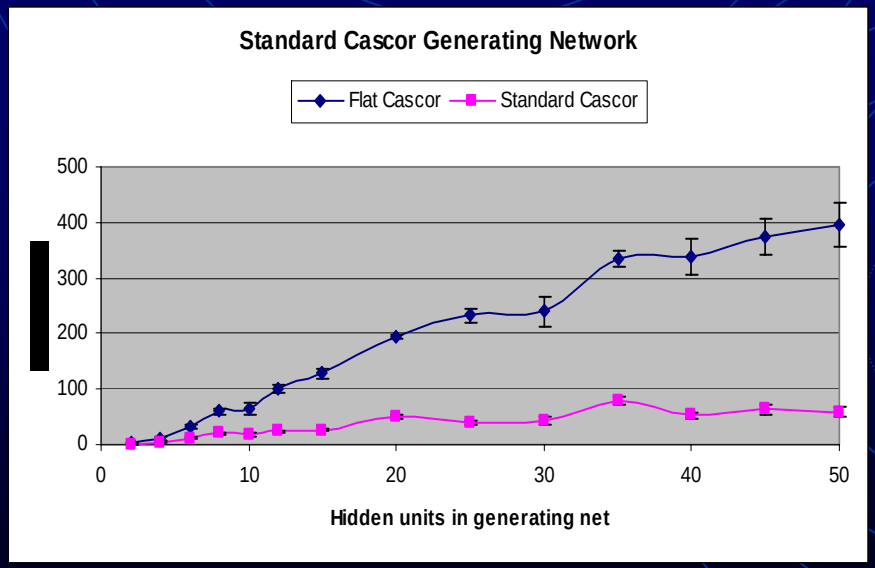
- Train flat and standard networks using those data

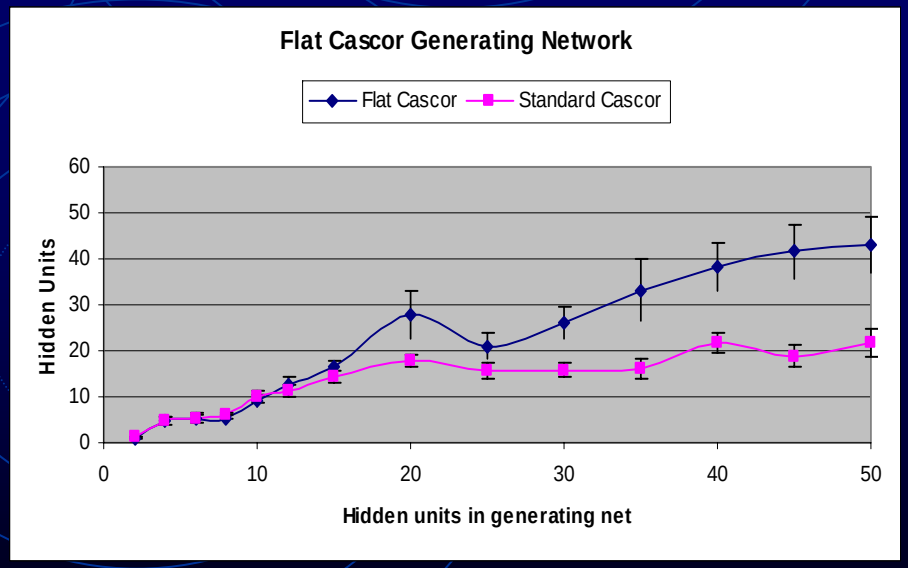
Performance Evaluation

- Recruited hidden units
- Computational cost for training the network
- Number of connection weights after training = Recall cost
- Accuracy = Error on train set
- Generalization = Error on test set (unseen during training)

Results and Discussion

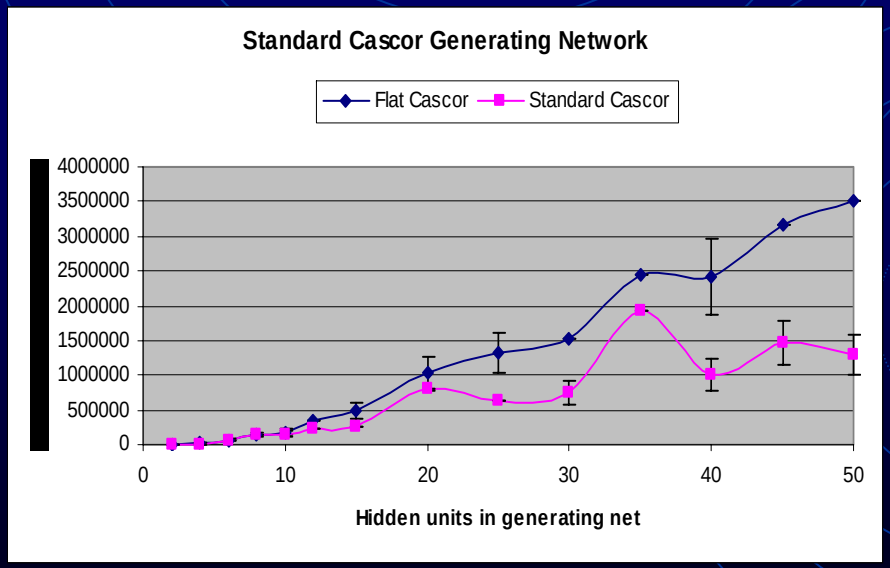
- *Lower values are better on results graphs*
- Recruited units count
  - Best: Standard Cascade-Correlation

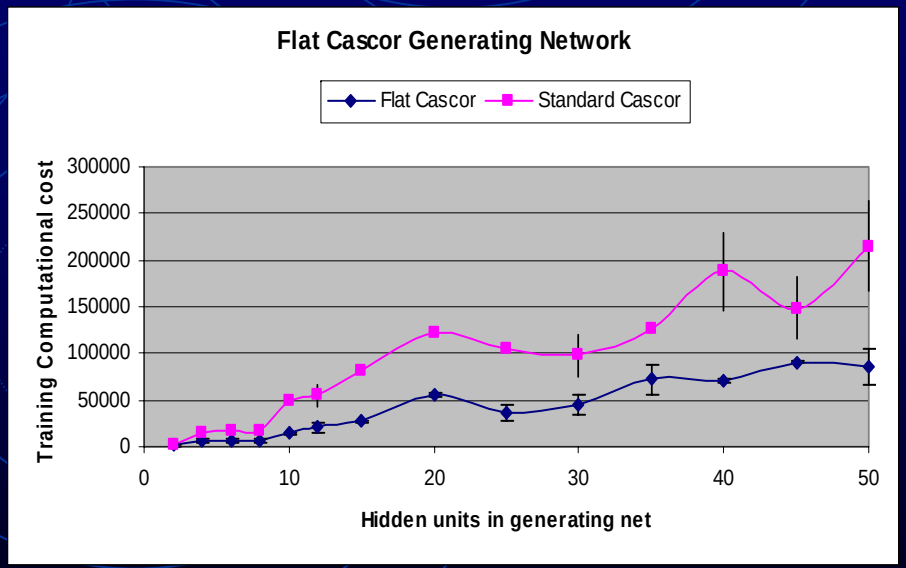




Results and Discussion

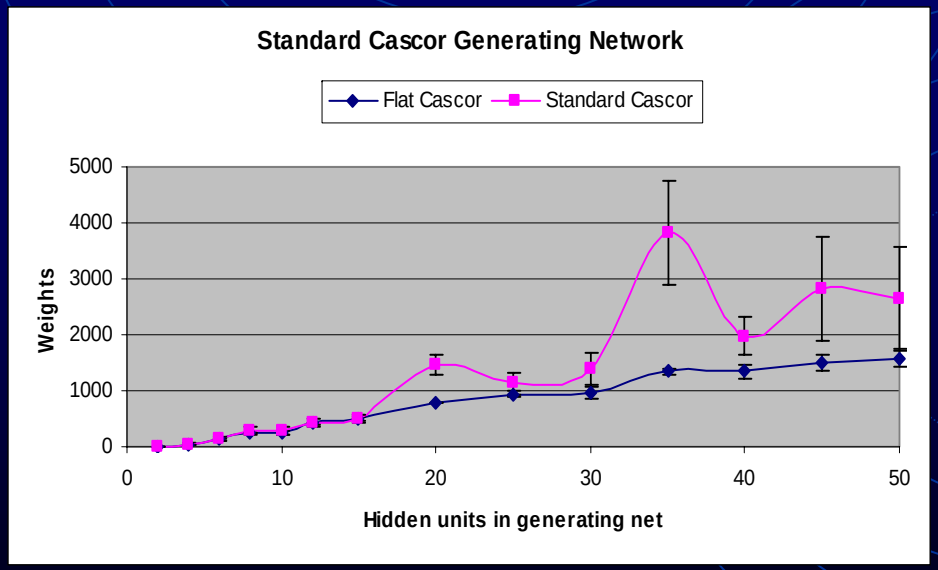
- Computational cost for training
  - Best: Network type matching the generating type

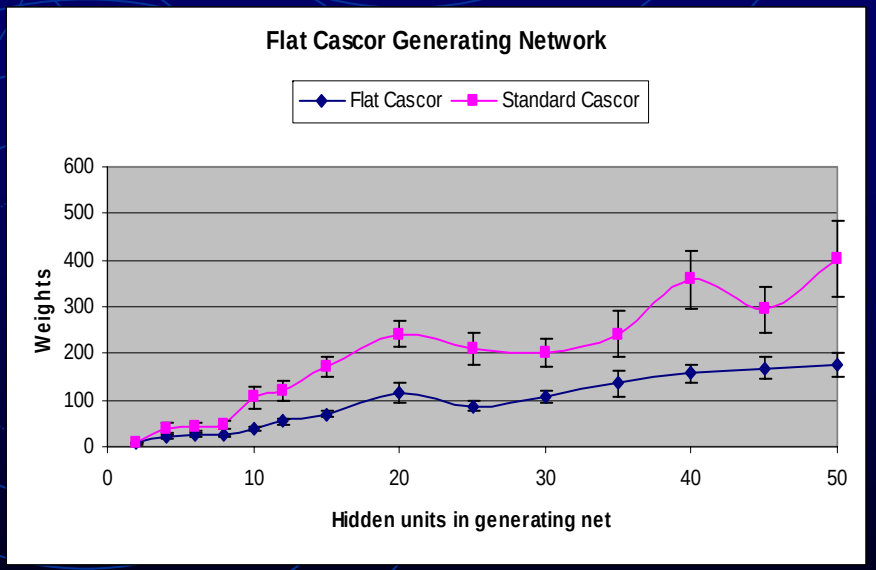




Results and Discussion

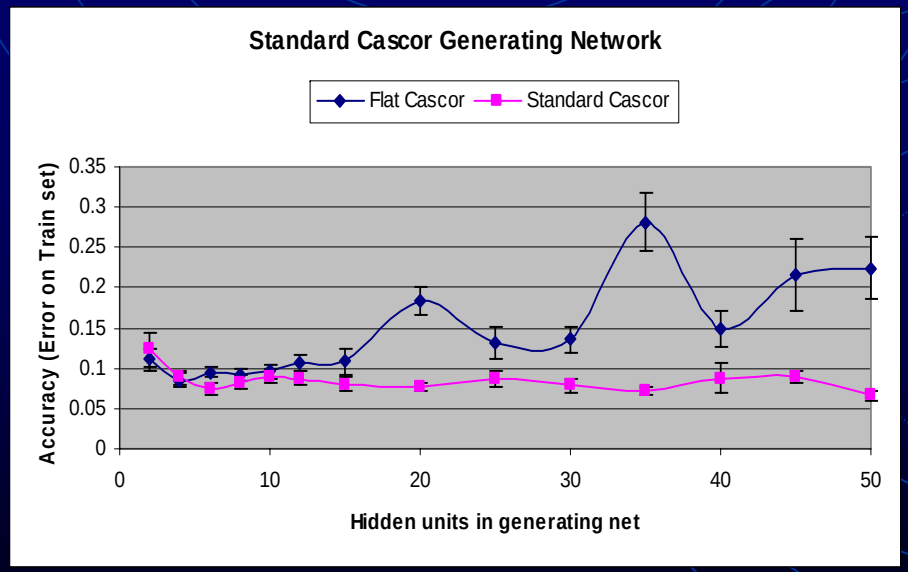
- Number of weights after training
  - Best: Flat Cascade-Correlation

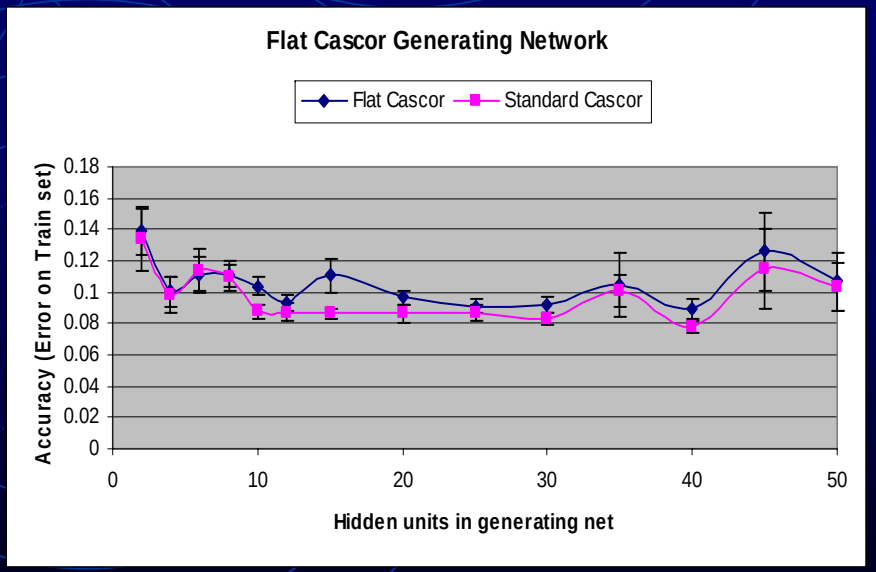




Results and Discussion

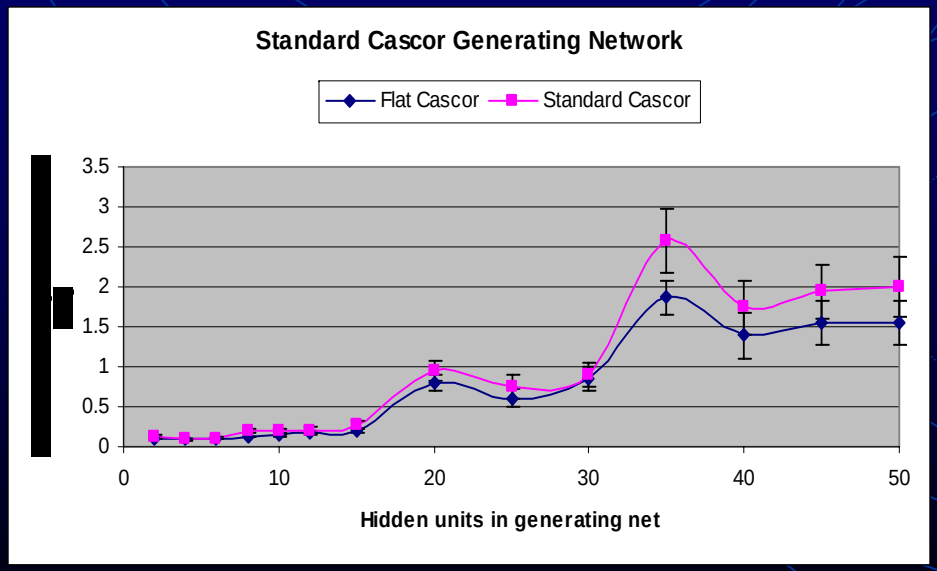
- Accuracy (Error on Train Set)
  - Best: Standard Cascade-Correlation

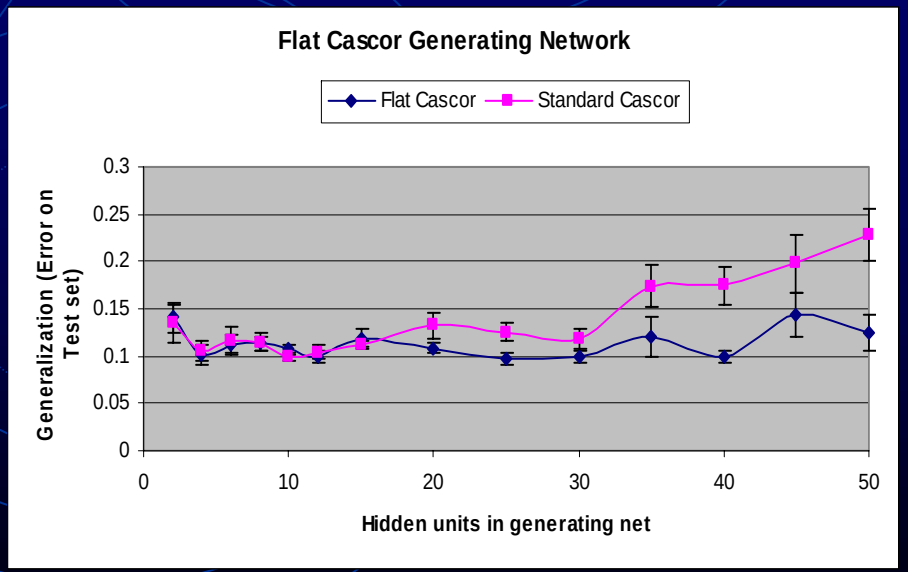




Results and Discussion

- Generalization (Error on Test Set)
  - Best: Flat
  - Consistent with Prechelt





Conclusion

- Both flat and standard networks were able to learn functions generated by both standard and flat-generating networks.
- For small complexity output functions (i.e. few hidden units in generating networks), performance differences are not significant.

Conclusion

- For large complexity output functions, trained networks of the same type tend to offer the best overall performance.
  - Asymmetrical difference
  - Flat networks approximate flat networks ***slightly*** better
  - Standard nets best approximate standard nets ***much*** better

Conclusion

- Which is best?
  - Generalization – Flat Cascade-Correlation
  - Otherwise, standard networks approximate flat better than the opposite.
- Comprehensively, standard networks perform better

Future Directions

- Sibling-Descendent Cascade-Correlation