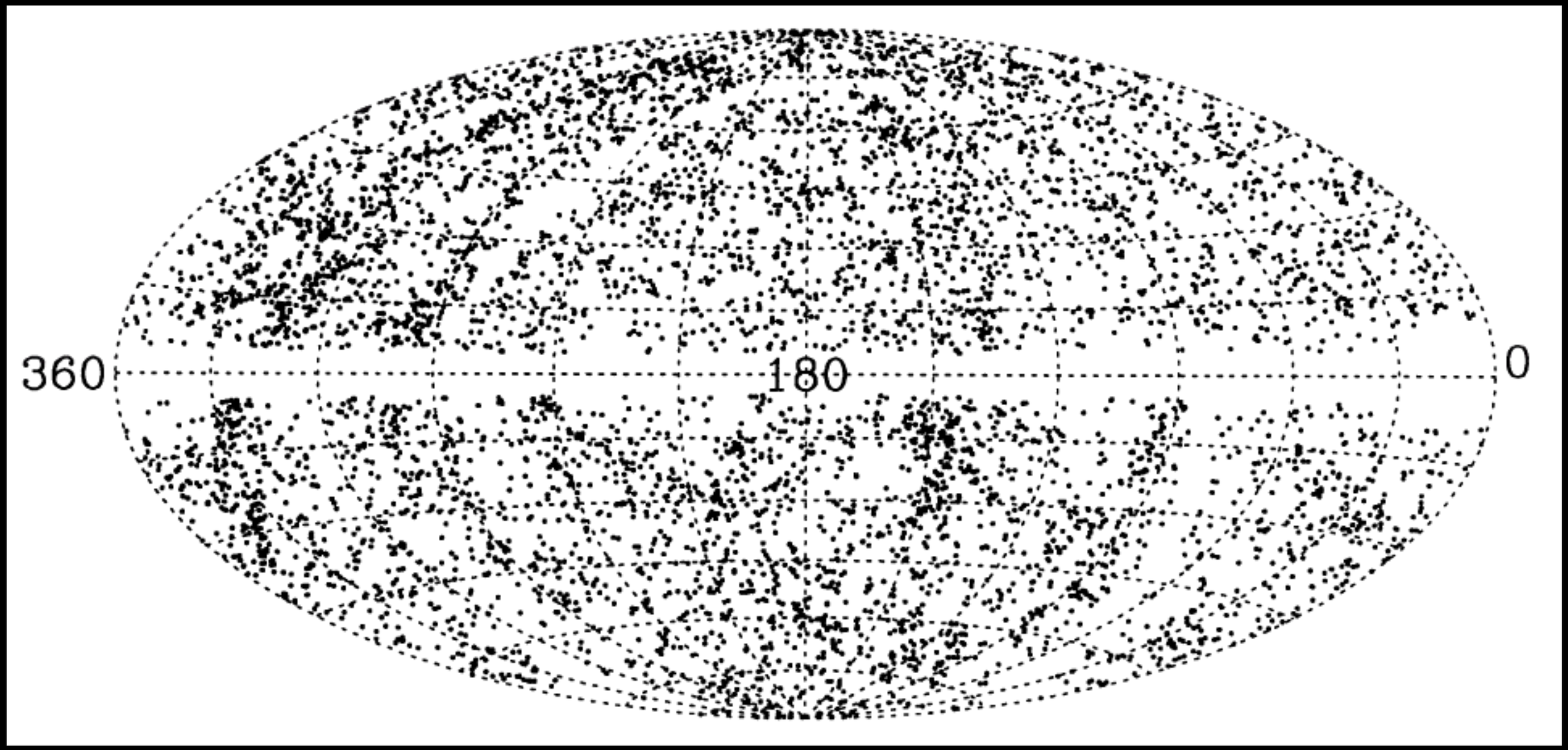
NEURAL NETS AND 1990S COSMOLOGY

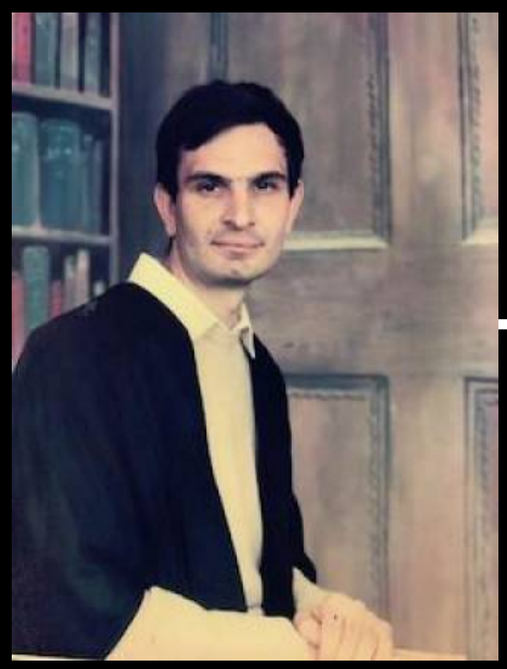
Caleb Scharf, Columbia University



1993

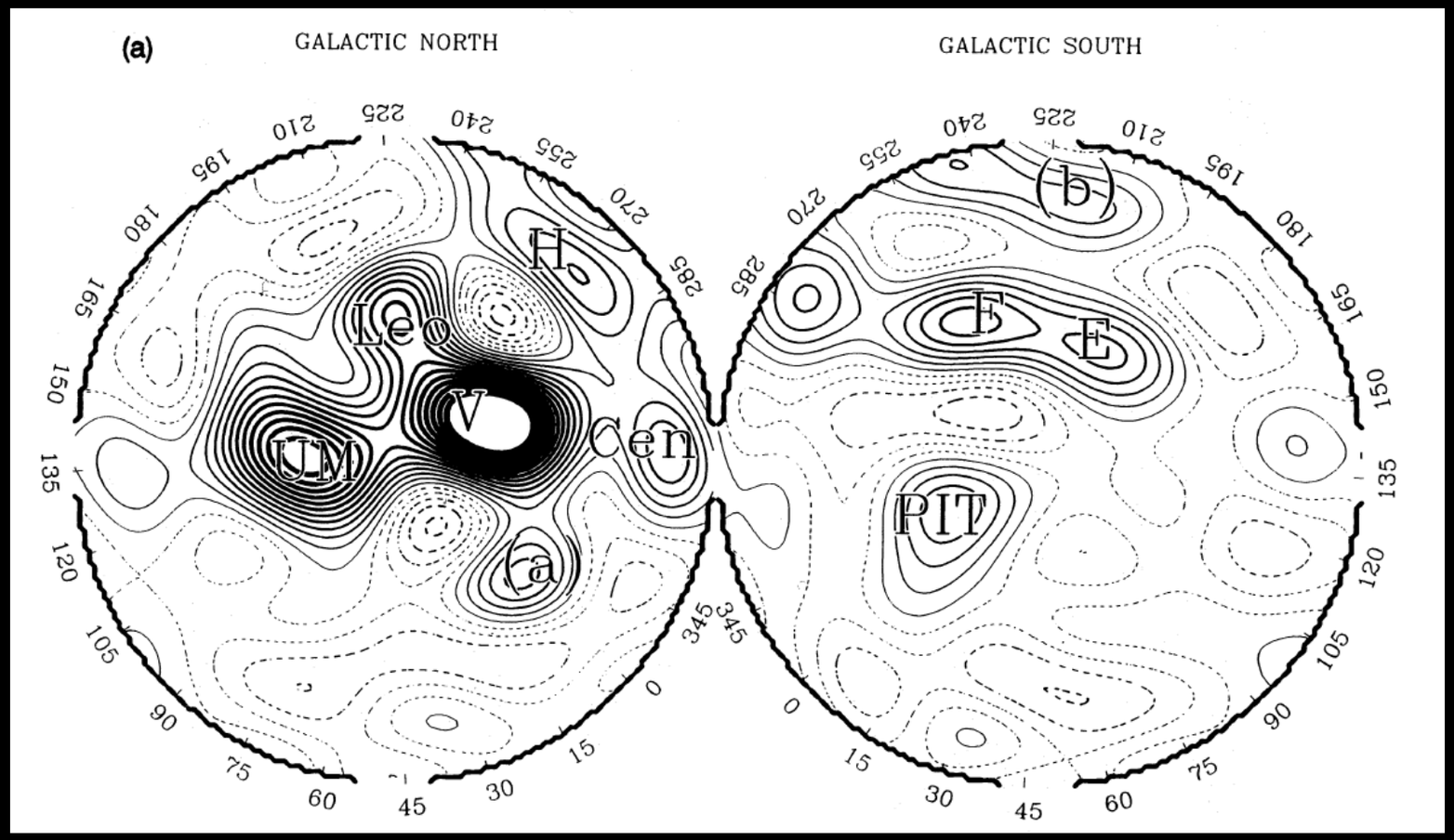
Evidence was growing for a low density cosmology

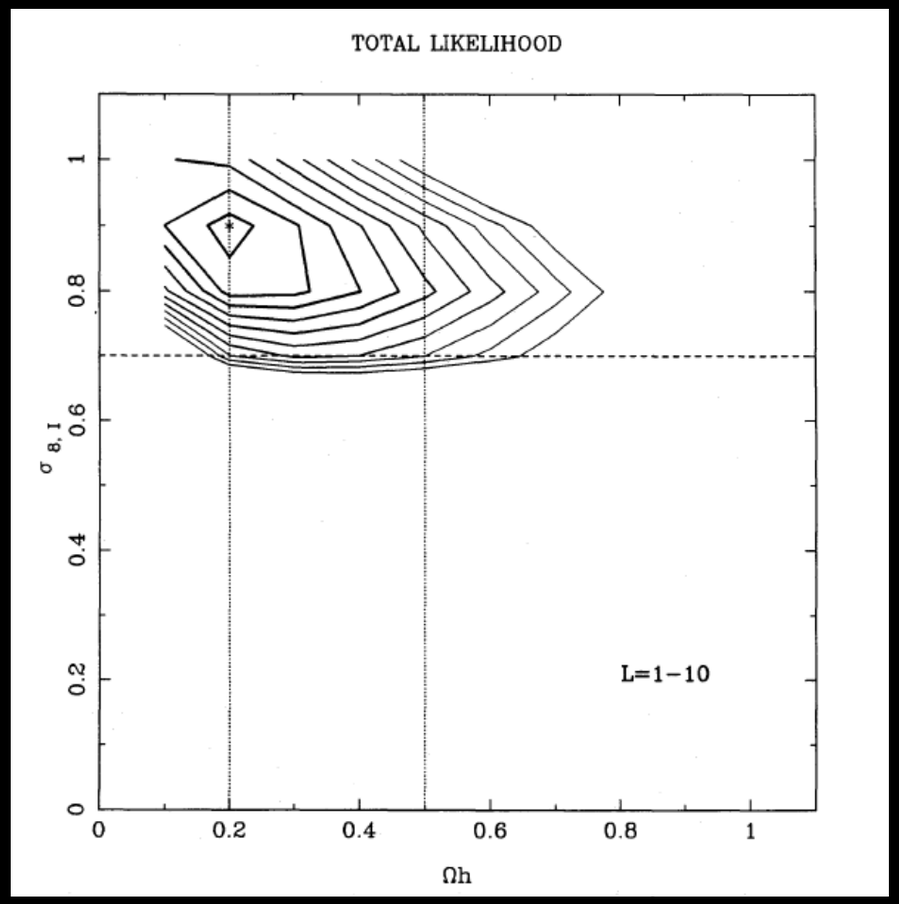
Applying increasingly sophisticated statistical analyses to galaxy surveys to pin down the power-spectrum of matter fluctuations was a popular industry



What about using this cool-sounding thing that was an Artificial Neural Network?

We had been using spherical harmonic analyses to study the angular power spectrum of ‘large’ galaxy surveys like the IRAS 0.7 Jy and 1.2 Jy catalogs - containing no less than 8,000 and 5,313 galaxies!





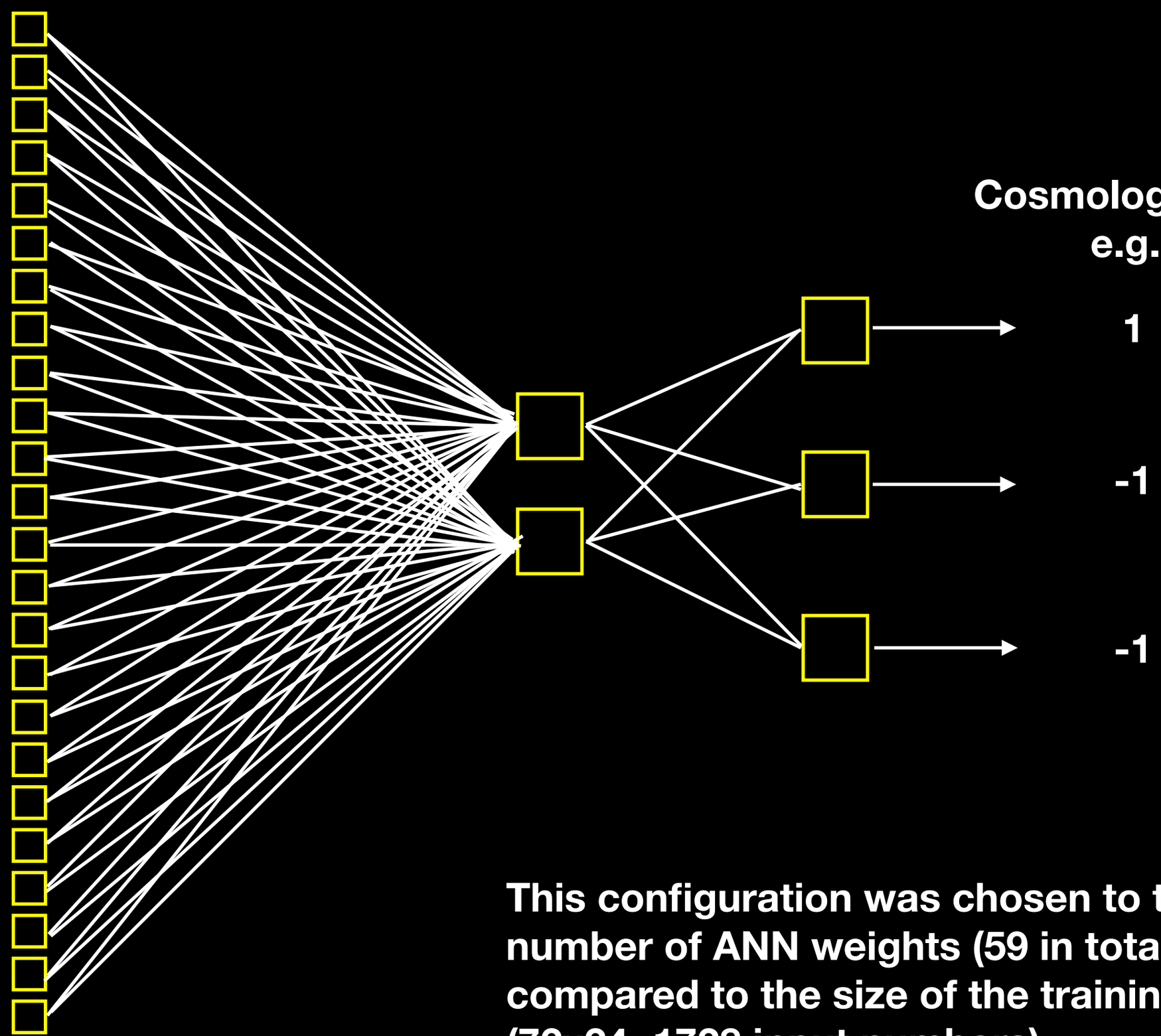
Scharf & Lahav 1993

But was there a way to make FULL use of both amplitude and phase information in the survey data?

Could we simply ‘ask’ a machine what kind of universe we lived in?

Single hidden layer, Backpropagation ANN

24 inputs (first 4 harmonics $l=1\text{-}4$), 2 nodes in hidden-later, 3 outputs



This configuration was chosen to try to keep the number of ANN weights (59 in total) smaller compared to the size of the training datasets (72x24=1728 input numbers)

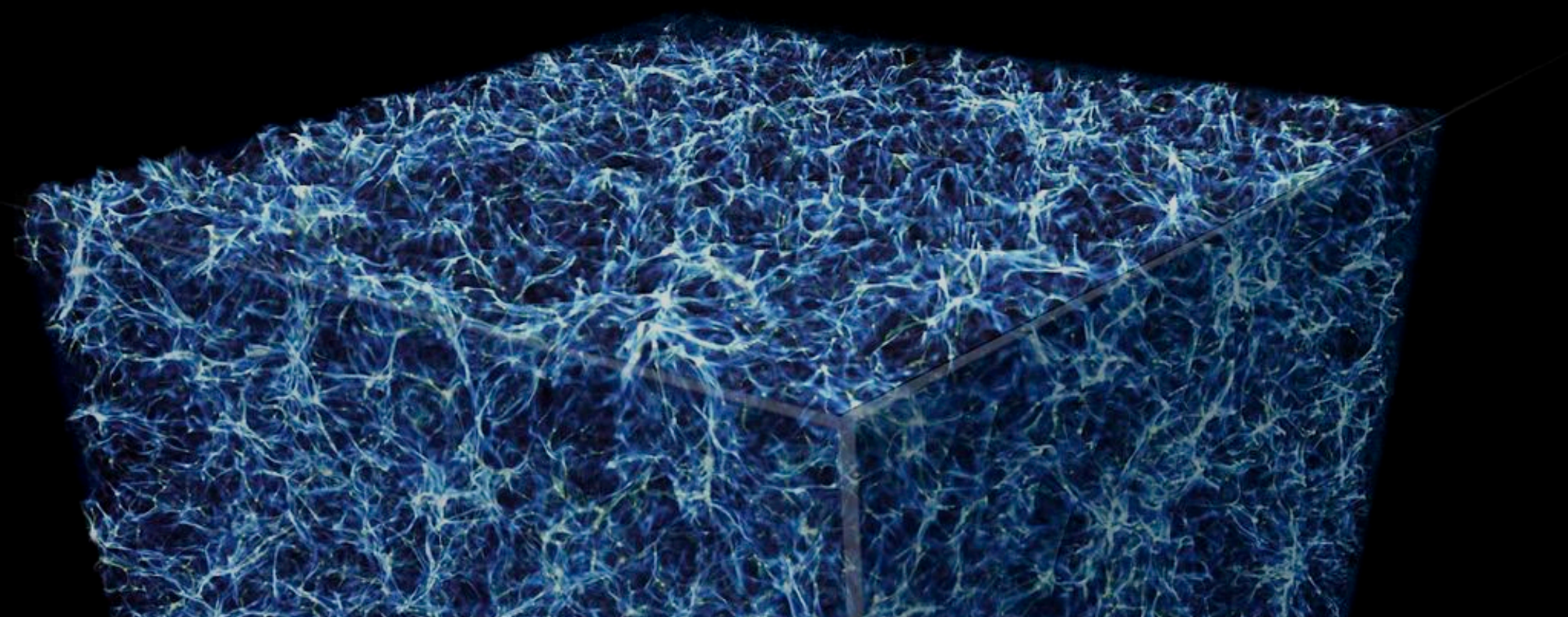
Training cosmologies

(1) 'Standard' CDM (i.e. power spectrum 'damped' with turnover *k*-space scale around Ωh^2) was considered to be $\Omega = 1 \quad h = 0.5$

(2) A scale-free spectrum ($P(k) \propto k^{-1}$)

C. Park's P^{3}M code, 1x10^6 to 1x10^7 particles in ~400 to 600 h^{-1}Mpc side boxes, then mock IRAS catalogs extracted of ~10,000 'galaxies'

(3) A Poisson model (literally random galaxies on the sky)

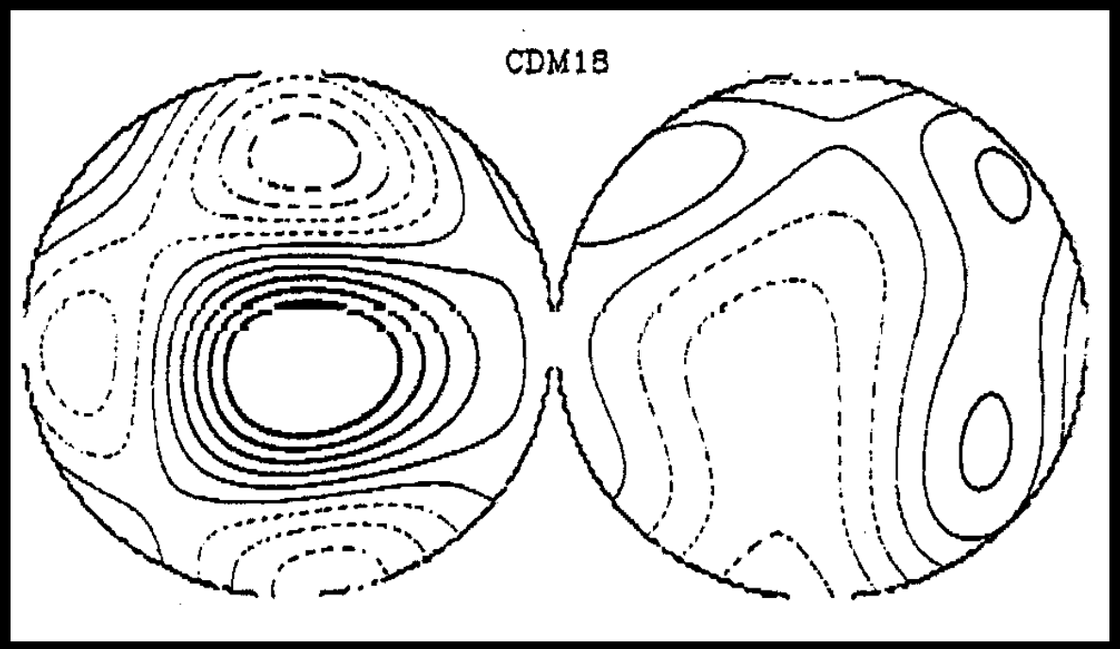


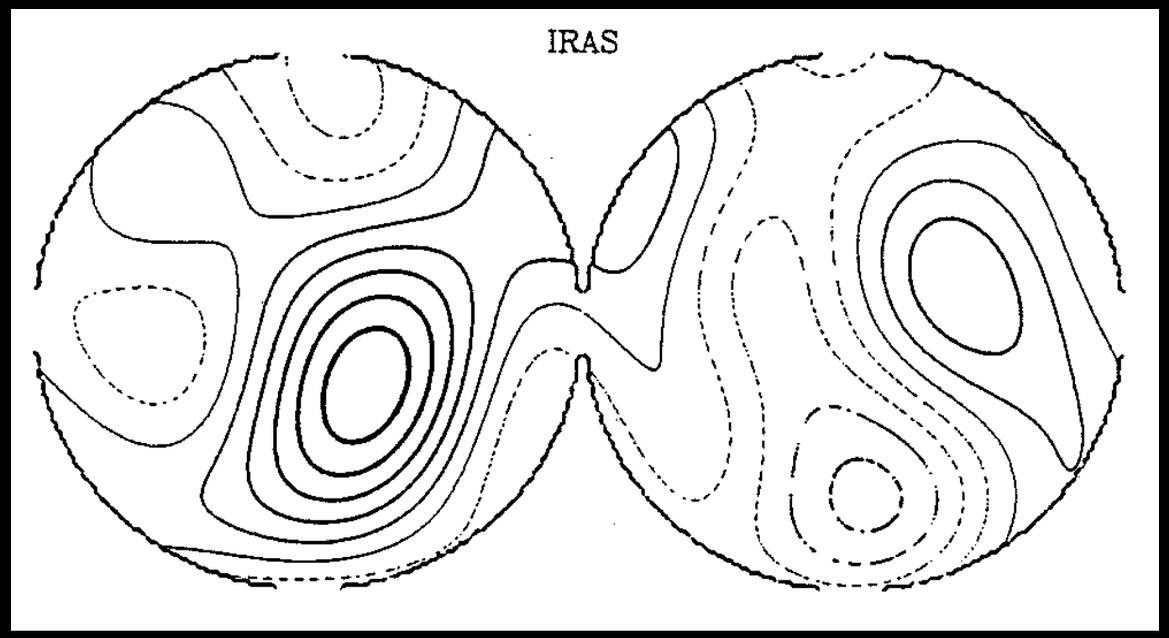
After training on 72 model realizations the ANN classified the true IRAS observations as belonging to the CDM class with 100% confidence, and 0% for other models.

This was contrary even to less-than-well-balanced priors…

What happened?

By chance, one of the CDM simulated IRAS surveys was a particularly good match to the real catalog, and the ANN recognized this





A cautionary tale - relevant also to current concerns over reproducibility in machine-learning applications to scientific data…

The same exquisite discriminatory skills that make deep learning algorithms so powerful - encoding subtle and complex correlations - can also be problematic

To try to mitigate this issue we Bootstrapped our training sets - 10 sets of 72 drawn randomly from the 72 available realizations.

This helped, and the final ‘answer’ could then be given as an average probability of the class of the REAL DATA over the 10 Bootstraps

CDM	k^{-1}	Poisson
27%	45%	17%

Using JUST dipole and quadrupole coefficients CDM was preferred at 95% (5% for k^{-1} and 0% for Poisson)

In testing the trained ANN on simulations we did note that the k^{-1} models were most likely to be misclassified…

A LESSON FROM THE 1990S

We showed that a learning system (shallow learning in this case) can be strongly biased due to the chance similarity between one statistical realization of a theoretical prediction and the real data. In a real sense the system 'P-hacks'.

In this case it was readily traceable, and the ANN's output was so decisive it raised a flag

But we can easily imagine situations where this kind of effect is more insidious...

The end results were only as good as the form of the chosen input data would allow - were spherical harmonic components really the best tool? Maybe not...but what would be?