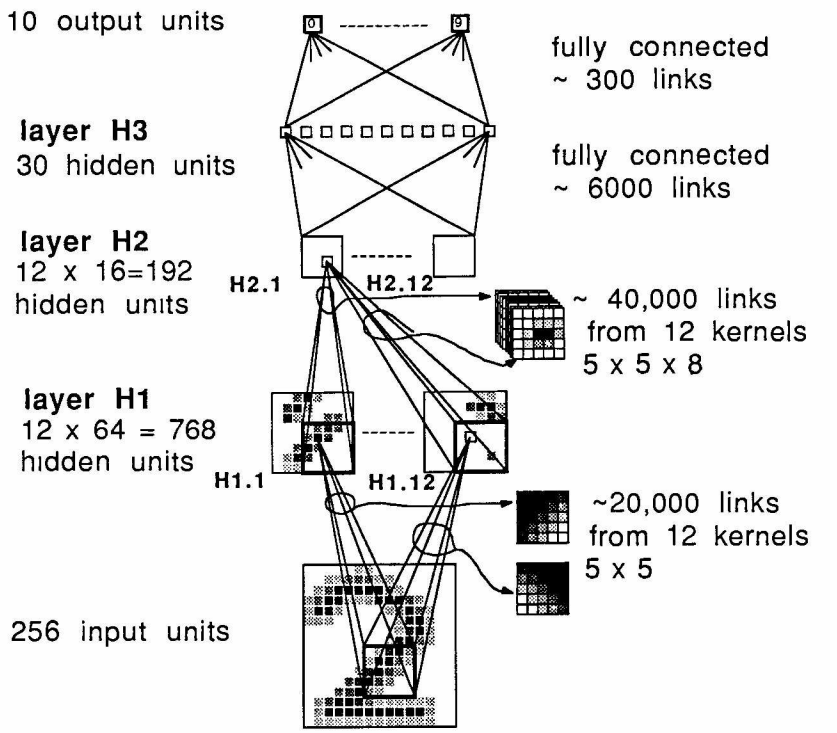

Figure 3 Log mean squared error (MSE) (top) and raw error rate (bottom) versus number of training passes

training set, 8 1% misclassifications on the test set, and 19 4% rejections for 1% error rate on the remaining test patterns. A full comparative study will be described in another paper

5.1 Comparison with Other Work. The first several stages of processing in our previous system (described in Denker *et al* 1989) involved convolutions in which the coefficients had been laboriously hand designed. In the present system, the first two layers of the network are constrained to be convolutional, but the system automatically learns the coefficients that make up the kernels. This "constrained backpropagation" is the key to success of the present system: it not only builds in shift-invariance, but vastly reduces the entropy, the Vapnik–Chervonenkis dimensionality, and the number of free parameters, thereby proportionately reducing the amount of training data required to achieve a given level