LRP 397/90 July 1990

Fast Non-Linear Extraction of Plasma Equilibrium Parameters Using a Neural Network Mapping

J.B. Lister and H. Schnurrenberger

C R P P

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE - SUISSE

FAST NON-LINEAR EXTRACTION OF PLASMA EQUILIBRIUM PARAMETERS USING A NEURAL NETWORK MAPPING

J.B. LISTER, H. SCHNURRENBERGER
Centre de Recherches en Physique des Plasmas,
Association Euratom - Confédération Suisse,
Ecole Polytechnique Fédérale de Lausanne,
1015 Lausanne, Switzerland

ABSTRACT The shaping of non-circular plasmas requires a non-linear mapping between the measured diagnostic signals and selected equilibrium parameters. The particular configuration of Neural Network known as the multi-layer perceptron provides a powerful and general technique for formulating an arbitrary continous non-linear multi-dimensional mapping. This technique has been successfully applied to the extraction of equilibrium parameters from measurements of single-null diverted plasmas in the DIII-D tokamak; the results are compared with a purely linear mapping. The method is promising, and hardware implementation is straightforward.

1. INTRODUCTION

The fast extraction of a reduced number of tokamak equilibrium parameters from the experimental data is essential for interpreting many diagnostic results as well as for real-time shaping of the plasma itself. In order to perform this extraction, we must develop a representation of the mapping between a selected measurement data vector $\underline{\mathbf{M}}$ and the extracted parameter vector $\underline{\mathbf{G}}$. We shall denote this mapping by G, such that $\underline{\mathbf{G}} = G(\underline{\mathbf{M}})$. The equilibrium problem is typical of many data interpretation problems, for which we posess a full physical description of the inverse mapping G^{-1}, in our case the Grad-Shafranov equation embodied in an axi-symmetric equilibrium code. Since we cannot derive a closed form of the mapping G from the physical model, we are required to construct an approximation of G, denoted by $\hat{G}$, which we hope will be valid over a given range of interest of the vector $\underline{\mathbf{M}}$.

In the case of almost centred circular plasmas, the mapping G is obtained by linearising, and then inverting, the known physical mapping G^{-1}. As the shape of the plasma varies and the range of equilibrium parameters increases, the linearised mapping becomes less and less reliable, especially in the extreme ranges of plasma parameters which are of most interest, or for particular parameters such as the X-point position which are less well linearised.

So far there have been three main approaches to approximating the non-linear mapping G . As the aim of this paper is to discuss the applicability of a proposed fourth method, we briefly discuss these three previous approaches in turn for comparison.

The first approach is to develop a description of G using ad hoc trial functions, and is successfully used on the DIII-D tokamak [1-3]. A database of reconstructed equilibria provides many examples of both measurement ($\underline{\mathbf{M}}$) and parameter ($\underline{\mathbf{G}}$) data. Trial functions are proposed, intuitively or by inspection, and the several free parameters in each of the trial functions are fitted by regression analysis. The starting point for the trial functions is a linearisation about a chosen operating point, based on a first order Taylor series expansion of the poloidal flux. The result is a set of different non-linear relationships $\hat{G}$ ($\underline{\mathbf{M}}$), valid and optimised for particular ranges of equilibria. These functions are then hardwired for real-time plasma control.

The second approach is to reduce the data by Principal Component Analysis (PCA) and subsequently develop a quadratic expansion for $\hat{G}$, which has been developed on ASDEX data [4] for use on the ASDEX-Upgrade tokamak [5]. Again, a well populated equilibrium database is used. A linear extraction of a significantly reduced number of 'principal components' is performed, and a quadratic relationship is then assumed:

$$\hat{\mathrm{G}}(\underline{\mathbf{M}}) = \underline{\mathbf{A}} + \underline{\mathbf{B}} \cdot \underline{\mathbf{M}}^* + \underline{\mathbf{M}}^{*\mathrm{T}} \cdot \underline{\mathbf{C}} \cdot \underline{\mathbf{M}}^* \quad ; \quad \underline{\mathbf{M}}^* = \underline{\mathbf{PCA}} \cdot \underline{\mathbf{M}} \tag{1}$$

where $\underline{\mathbf{PCA}}$ is the principal component projection matrix and the matrices $\underline{\mathbf{A}}$, $\underline{\mathbf{B}}$, $\underline{\mathbf{C}}$ together provide a non-linear mapping. We are restricted to a small dimension of $\underline{\mathbf{M}}^*$ in order that the large matrix $\underline{\mathbf{C}}$ be manageable, and realistically implementable in hardware. One significant advantage over the method of ad hoc trial functions is the generality of the expression (1), allowing its straightforward implementation in hardware [6].

The third method presupposes no functional form of G except that it be locally linearisable over a significant distance in the space of $\underline{\mathbf{G}}$. The linearised matrix $\hat{G}$

must be varied as the space of $\underline{\mathbf{G}}$ is explored by the plasma evolution. Each local matrix $\hat{G}$ can either be deduced by linearising a subset of an equilibrium database, or it can be deduced directly from the equilibrium [7]. The former approach fits a hyperplane through a portion of the function G, minimising the mean square residuals of the set of examples of $\underline{\mathbf{G}}$. The latter approach calculates the local tangential hyperplane. These linear methods have been developed for the control of the TCV and Alcator C-MOD tokamak plasmas [8].

These three methods all have both advantages and disadvantages, which are summarised in Table I, in which the questions of hardware implementation are left out. In the remainder of this paper, we investigate the applicability of a new technique. We introduce the multi-layer perceptron in Section 2, study its performance on non-centred circular plasmas in Section 3 and on DIII-D tokamak equilibria in Section 4, and discuss the results in Section 5.

2. THE MULTI LAYER PERCEPTRON

The Multi Layer Perceptron (MLP) is one of the fruits of an expanding interest in the Neural Network approach to Artificial Intelligence. The field has an oblique and mystifying vocabulary which we deliberately avoid; the interested reader is referred to Refs [9, 10] for further background information.

The MLP is an explicit non-linear and continuous mathematical relationship between a multi-variable input data vector (in our case diagnostic information) and a multi-variable output data vector (in our case reduced equilibrium parameters). The MLP is often represented schematically as in Fig.1. The input vector, of dimension N_1, is linearly projected by a matrix $\underline{\mathbf{W}}_{12}$ onto a usually reduced intermediate vector of dimension N_2. This operation is reminiscent of the initial PCA projection onto a reduced dimension vector $\mathbf{M}^*$. As in the case of PCA, the size of this intermediate vector is the only design choice when specifying the MLP configuration. On the other hand, we no longer have to handle the large and probably sparse matrix $\underline{\mathbf{C}}$ ($N_3 \times N_2 \times N_2$). The reduced vector is then passed element by element through a compression function S, known as the sigmoidal function, which is bounded ± 1, continuous, differentiable and monotonically increasing. The chosen sigmoidal function is:

$$S(x) = \frac{2}{(1 + e^{-x})} - 1 \qquad (2)$$

The compressed vector is then linearly transformed by a second matrix $\underline{\mathbf{W}}_{23}$ into an output vector of dimension N_3. Such a configuration is referred to as a 1-hidden-layer MLP (an MLP-1); if we recompress and re-transform the output vector to produce a new output vector, we have a 2-hidden-layer MLP (an MLP-2) and so on. If we make a simple linear projection we have the historically special case of an MLP-0.

The MLP-1 representation of $\hat{G}$ can therefore be written explicitly as

$$\hat{G}\ (\mathbf{M}) = \underline{\mathbf{W}}_{23} \bullet \left[\frac{2}{(1 + e^{-\underline{\mathbf{W}}_{12} \cdot \mathbf{M}})} - 1\right] \qquad (3)$$

The vector exponent denotes an element by element exponentiation, giving a vector result. In practice, a constant term is added to each *non*-output layer to provide an offset bias; this and other details of the practical implementation of an MLP are discussed in Ref. [11].

A wide range of literature can be found on the generality of the MLP-1 and MLP-2 functions. It has been demonstrated that all bounded continuous functions can be expressed as an MLP-2 over a given volume in the space of $\mathbf{M}$ [12]. A similar demonstration has been proposed for the case of the MLP-1 [13, 14]. The choice between MLP-1 and MLP-2 appears to be dictated by whether the simpler MLP-1 is adequate for a given functional mapping [11]. The work in this paper is restricted to MLP-1 mappings which are shown to be adequate for the particular problem.

Having established the applicability of an MLP-1 network, two problems remain:

(i) choosing the size of the condensed hidden vector and

(ii) finding the optimal linear transformation matrices $\underline{\mathbf{W}}_{12}$ and $\underline{\mathbf{W}}_{23}$.

The first question is a compromise. Too few elements in the intermediate vector result in a too encoded form of the output, with linearly interdependent output variables. Too many elements lead to a badly defined set of matrices, and an inefficient solution in terms of hardware. As the number of free parameters increases, we also risk finding a solution which is adapted to the fitted examples, rather than generalising the functional dependence, the well-known problem of overfitting or bias.

The second question is fully discussed in Ref. [11]. The aim is to minimise the sum of the squares of the errors between examples of data which have been modelled, $\underline{G}_i = G(\underline{M}_i)$, and the estimated form of $\hat{G}$ given by the MLP-1 mapping (3). We define an RMS residual for each output denoted j:

$$\sigma_j = \frac{1}{2}\sqrt{\frac{\sum_i (\underline{G}i_j - \hat{G}(\underline{M}i)_j)^2}{\text{number of examples}}} \qquad (4)$$

In order to restrict the range of the values of the matrices $\underline{W}_{12}$ and $\underline{W}_{23}$, we normalise the vectors $\underline{G}$ and $\underline{M}$ to lie within the cubes $[-1, 1]^{N_3}$ and $[-1, 1]^{N_1}$ respectively. Expression (4) for σ_j therefore corresponds to the residual as a fraction of the full-scale range of the j-th parameter (% FS).

In practice, we have minimised the RMS value of the σ_j's, using only half of the available mapping examples $(\underline{M}_i, \underline{G}_i)$. The remaining examples are used to test the general validity of the approximated function. The minimisation is performed using a modification of the standard Gradient Descent Technique [11].

3. CIRCULAR PLASMAS

In this section we consider the testcase of an unshaped plasma immerged in a purely vertical poloidal field, within a 3:1 rectangular limiter aperture, Fig. 2. In the large aspect ratio approximation, the poloidal flux distribution is given by

$$\psi(R,Z,\Lambda) = \frac{\mu_0 I_p R}{2\pi}\left[\ln 8R/\rho - 2\right] - \frac{\mu_0 I_p}{4\pi}\left[\ln(\rho/a) + (\Lambda + 0.5)(1-a^2/\rho^2)\right]\rho\cos\omega' \qquad (5)$$

where ρ, a, ω' are shown in the figure.

The free variables are the major radius (R), vertical position (Z) and poloidal asymmetry factor (Λ). The plasma current I_p was held constant.

The detection signals were chosen to be a set of poloidal flux loops placed slightly outside the limiter aperture, shown in the figure, and the mapping inputs were the differences in flux with respect to one of the loops. This testcase has the required property of being non-linear by virtue of Eq. (5), as well as being simple and descriptive of the full problem.

A set of 1000 equilibria was generated, of which 500 were to be mapped, and 500 were used for testing the quality of the derived mapping. The values of [R, Z, Λ] were varied randomly and uniformly between [0.8 - 0.9, -0.48 - 0.48, 0.5 - 4.0] respectively. The first 500 examples were mapped by both a linear relationship and by the MLP-1 with N_2 = 10. The quality of the representation was taken to be the mean square residual, normalised to the full range of each parameter. This exercise was carried out for a varying number of flux-loops, obtaining the results shown in Fig. 3.

The linear representation of $\psi \rightarrow [R,Z,\Lambda]$ was good for very large numbers of flux-loops. However, for smaller numbers of loops, fewer than 15, the representation

was inadequate. The MLP-1 mapping maintained low residuals even when the total number of flux loops was decreased to as few as 6. An example of the quality of representation, fitted value vs. actual value, is shown in Fig. 4 for this case with only 6 input signals and 5 flux-differences.

The MLP-1 clearly is able to use its inherent non-linearity to solve the type of non-linear mapping problem represented by Eq. (5). Only the plasma vertical position is well represented by the linear mapping, the major radius and asymmetry factor being essentially non-linearly related to the fluxes.

4. APPLICATION TO DIII-D TOKAMAK EXPERIMENTAL DATA

A DIII-D tokamak database of experimentally achieved single-null diverted discharge equilibria was also used to test the applicability of the method. Since the equilibria are experimental, both $\underline{\mathbf{M}}$ and $\underline{\mathbf{G}}$ have intrinsic noise, the latter being reconstructed from an inverse-equilibrium code [15]. During the MLP fitting, we again used a random set of 500 equilibria as the training set, and the quality of the fit was subsequently tested on 500 different equilibria.

Firstly, we must select the vectors $\underline{\mathbf{G}}$ and $\underline{\mathbf{M}}$ to be mapped. Our aim is to extract a certain number of geometric parameters, not necessarily significant in the formal sense of the dynamic control problem, but important for plasma operation. Typical parameters and their possible operational importance are: X-point position, heat load and configuration; Outer gap, antenna-plasma spacing; Top gap, elongation control; Current centroid, fast vertical feedback; Inner gap, limiter-separation; Core-elongation, vertical stability; Internal inductance, dynamics of vertical control; β, heating control; area, current control; q_{95}, current control. This list can be extended at will, provided that the tokamak itself can independently vary these parameters.

The first seven parameters have been conventionally used by DIII-D for shape control [3] and were used as the vector $\underline{\mathbf{G}}$ in a first test. For input variables we chose: (i) a set of tangential magnetic field probes (22 coils); (ii) a set of magnetic flux loop flux-differences with respect to the mid-plane inside wall loop (20 loops); (iii) both field probes and flux loops. Figure 5 shows a schematic layout of the measurements and of the derived parameters.

We have a further choice of mappings depending on the normalisation with respect to the plasma current, I_p. If we map $[\psi, B_{pol}] \rightarrow I_p \bullet$ Geometry, the dominant

dependence is the plasma current quasi-proportionality. By varying the range of I_p, we change the apparent full-scale precision of the representation. We therefore chose to map $[\psi/I_p, B_{pol}/I_p] \rightarrow$ Geometry to assess the precision of the method.

The cycle: choice of measurement - choice of output vector - choice of N_2 - optimisation of $\mathbb{W}_{12}$, $\mathbb{W}_{23}$ - test of RMS residues σ_j, was carried out several times on the same data set but for different configurations, with the results of Table II. In all cases we chose $N_2 = 10$. The RMS residuals obtained from a linearised fit are also indicated in the Table for comparison. If the input-output relationship is essentially linear, the two residuals will be the same. If the problem is non-linear, the linearised mapping will be poorer.

For each set of input data, the MLP-1 performed on average better than the linearised fit, showing that the MLP-1 was able to provide a useful degree of curvature in the approximate mapping $\hat{G}$. Combining the probes and flux-loops provided the best fit in both cases as expected. The tangential field probes alone provide a better definition of the plasma geometry than the flux-loops alone. We interpret this in terms of a better 'anchorage' of the separatrix of these single-null diverted plasmas where it crosses the vessel.

The contributions of the different σ_j to $< \sigma_j^2 >$ are very different. In the linearised case, the output variables are fitted independently and weighting them differently does not change the solution. In the MLP-1 case, we can reduce $< \sigma_j^2 >$ by worsening the fit on one output variable to benefit any less well fitted variables. The σ_j were simply weighted inversely to their full-scale range using (3).

To push the technique further, the full exercise was repeated with 13 output variables, and the σ_j's are summarised in Table III. The linearised residuals are also shown, as well as the physical range of each parameter, and the minimum and

maximum absolute errors. For all parameters but one, the MLP-1 representation is significantly better than the linear mapping; the σ_j are generally a factor 1.5 - 2.0 smaller and the maximum errors of both signs are also significantly reduced. The parameter which is least well fitted is the top triangularity (σ_j = 5.8% FS). The MLP-1 mapping was clearly constrained to reduce this particular σ_j, and this was done at the expense of the vertical position of the plasma current (σ_j = 0.9% FS). Clearly such a global mapping would benefit from a selected output parameter weighting, such as the inverse residual of the linearised fit.

Figure 6 shows the residuals for four of the control parameters: inner gap, top gap, geometric centre and X-point Z-position. We compare the MLP-1 fit with the DIII-D adhoc function fit [3]. Using the MLP-1 mapping we generally obtain an improvement of 30-50 % in the residuals compared with a general linearisation using all of the 42 input parameters. The simple linearisation surprisingly also performed much better than the more complicated trial functions. We consider that this is partly due to the use of more input signals, and partly due to the removal of constraints on the coefficients used in [3].

As a further test of the advantages of the MLP-1 mapping, we compared the quality of the linear and MLP-1 mappings when the number of input signals was varied. The full set of 42 input variables was reduced stepwise, removing 5 randomly chosen inputs at a time. Each set of reduced input data was then mapped to the output data by both linear and MLP-1 mappings. Figure 7 shows the results of this test. The MLP-1 mapping led to significantly smaller mean-square residuals than the linear mapping, as in the test-case. The difference is, however, less marked than for the test case. This fact can be partly due to the inherent noise in the "real" output data, and partly due to a less significant exploration of the non-linear function to be mapped.

We briefly consider other mappings which may become interesting. Since the plasma current is extractable from the input data via $\mu o I_p = \int_c \underline{B} \bullet \underline{dl}$, a general mapping function should be able to map $\psi \rightarrow$ position, performing the division implicitly. This was tested, and $< \sigma_j^2 >$ was barely different from the case in which we mapped $\psi/I_p \rightarrow$ position. The linearised case was unusable, as to be expected. Although the un-normalised mapping is most relevant for the poloidal flux control, this un-normalised $\rightarrow$ normalised mapping would be necessary for, q_{95} control, β_p or li control.

Since q_{95} control is especially important when the shape is varying significantly during the discharge, we examined the mapping $[\psi/I_p, B_{pol}/I_p] \rightarrow B_\phi/(I_p \bullet q_{95})$, which is proportional to the ratio between q_{95} and the cylindrical estimate q_I. This value varied over a range of 1.9 - 3.1 for the database used (κ between 1.63 and 2.08), but will cover an even larger range in the TCV tokamak at high elongation. Using all the flux-loops and field probes, the MLP-1 mapping yielded σ_j = 0.8% FS, compared with σ_j = 1.3 % FS for the linear mapping. The improvement with the non-linear mapping is significant, and probably useful, especially at the higher elongation of the TCV tokamak. Reducing the input to just the field probes gave σ_j = 1.0%, compared with 1.7 % linearised; using a random set of only 11 probes gave σ_j = 2.7 % compared with 4.1 % linearised The same general improvement was found as with the shape parameters.

5. DISCUSSION

The non-linear mapping offered by an MLP-1 configuration of Neural Network provides a new approach to the inherently non-linear problem of the extraction of the plasma configuration from the plasma diagnostics. A simple testcase demonstrates the applicability of this class of mapping to the particular problem, especially when only a few input signals are used. The method provides a useful represent-ation of the DIII-D single-null diverted plasma equilibria, although the improvement over the linearized fit is only of the order of 50%. As tokamaks become more shaped, as will be the case for the TCV plasmas, we expect the improvement to become even more significant, as the mapping G becomes more non-linear. The method described in this paper provides the following potential advantages:

- A significant reduction in the dimensions of the intermediate data, as with PCA, but without the constraint of a subsequent quadratic expansion
- Normalisation can be implicit, explicitly hardwired, or mixed
- A general representation which does not require new hardware for each function to be mapped
- The hardware implementation is only a slight modification of the ASDEX-UG, TCV and ALCATOR C-MOD solutions
- No switching or much less switching of the mapping will be required, than is necessary for piecewise linear mappings
- More robust mapping in the presence of input noise or matrix noise [11].

The method also looks very promising for the mapping of more varied sets of equilibria; the data set used was restricted to single-null diverted discharges. We expect the mapping to be discontinuous in its derivative on transition between limiter - single-null and single-null - double-null configurations. The MLP mapping may well then represent a significant improvement over the mappings currently

being studied, all of which are planned to switch on such a transition. In such a case we are effectively using the MLP ability to perform a classification (single-null : double-null : limiter) as part of the mapping, recalling the Artificial Intelligence origins of the method.

ACKNOWLEDGEMENTS

We are most grateful to the DIII-D physics team, especially Drs. John Ferron and Lang Lao for access to the DIII-D database used in this study. We thank Drs. Tony Taylor and Chris Bishop for their comments on this paper. The applicability of the MLP to this problem was originally discussed with Dr. Robert W. Means. The work was partly funded by the Fonds national suisse de la recherche scientifique.

REFERENCES

[1] OSBORNE, T.H., FUKUMOTO, H., HOSOGANE, N. et al., Bull. Am. Phys. Soc. **31**, 9 (1986) 1502.

[2] FUKUMOTO, H., HOWL, W., JACKSON, G. et al., private communication - Shape Control Workshop, Lausanne, 1987.

[3] KINOSHITA, S., FUKUMOTO, H., KELLMAN, A.G., et al., General Atomics Report GA A 19584 (1989)

[4] BRAAMS, B.J., JILGE, W. and LACKNER, K., Nuclear Fusion **26** (1986) 699.

[5] BRUHNS, H. and MC. CARTHY, P., private communication - July 1989.

[6] JULICH, A. and RICHTER, H., private communication July 1989.

[7] HOFMANN, F. and TONETTI, G., Nuclear Fusion **28** (1988) 519.

[8] LISTER, J.B., MARMILLOD, PH., and MORET, J.M., Lausanne Report LRP 332/87.

[9] RUMELHART, D.E., and MC. CLELLAND, J.L., Parallel Distributed Processing: Explorations of the Microstructures of Cognition, Vols I, II and III, MIT Press (1986).

[10] LIPPMANN, IEEE ASSP Magazine **4** (1987) 4.

[11] LISTER, J.B.and SCHNURRENBERGER, H., Lausanne Report LRP 398/90

[12] CYBENKO, G., Tufts University Report, Medford, U.S.A. 1988.

[13] HECHT-NIELSEN, R., Proc. INNS Annual Meeting, San Diego, U.S.A., (1988).

[14] FUNAHASHI, K., Neural Networks **2** (1988) 183.

[15] LAO, J.L., ST. JOHN, H., STAMBAUGH, R.D., KELLMAN, A.G., PFEIFFER, W., Nuclear Fusion **25** (1985) 1611.

FIGURE CAPTIONS

1. The schematic description of a Multi Layer Perceptron.

2. Sketch of the test-case geometry. Variables are the major radius (R), vertical position (Z) and asymmetry factor (Λ)

3. Quality of representation with (a) linear mapping and (b) MLP-1 mapping, as the number of flux loops is varied

4. Quality of the mapping for 6 flux loop differences; (a) MLP-1 mapping and (b) linear mapping

5. A schematic of the DIII-D tokamak, showing the measurements and some of the derived plasma parameters.

6. Comparison between the trial function fit (upper) and MLP-1 representation(lower) for 4 control parameters (Units are cm.).

7. The Mean Square residuals, for: a) Random linear mapping and b) Random MLP mapping (asterisks).

TABLE TITLES

Table I

Comparison of three currently used methods

Table II

Residuals for different categories of input data, fitting with both an MLP-1 and linearised.

Table III

Residuals using an MLP-1 or linearised fitting. The input used both flux and field probes, the output is 13 geometrical parameters.

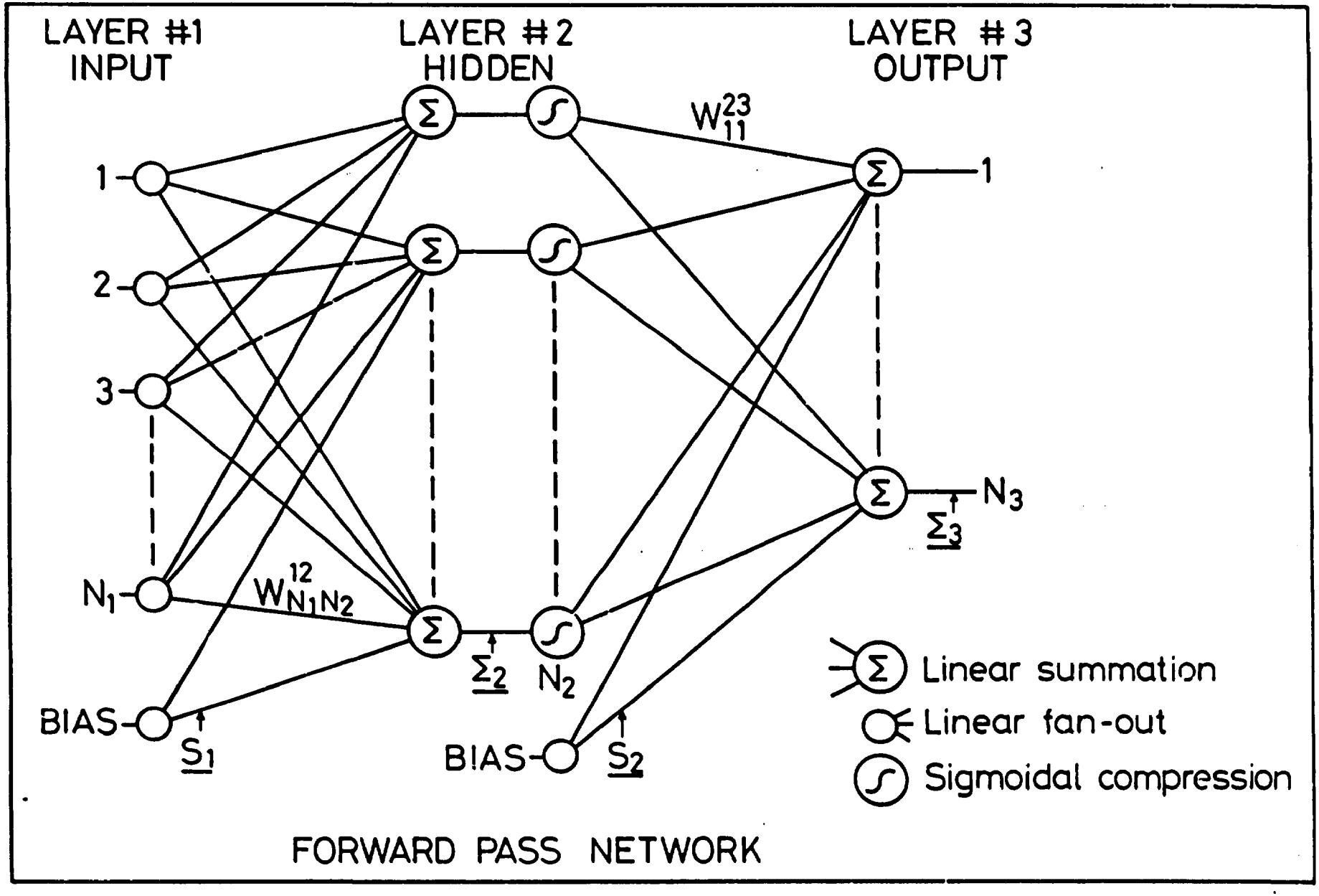
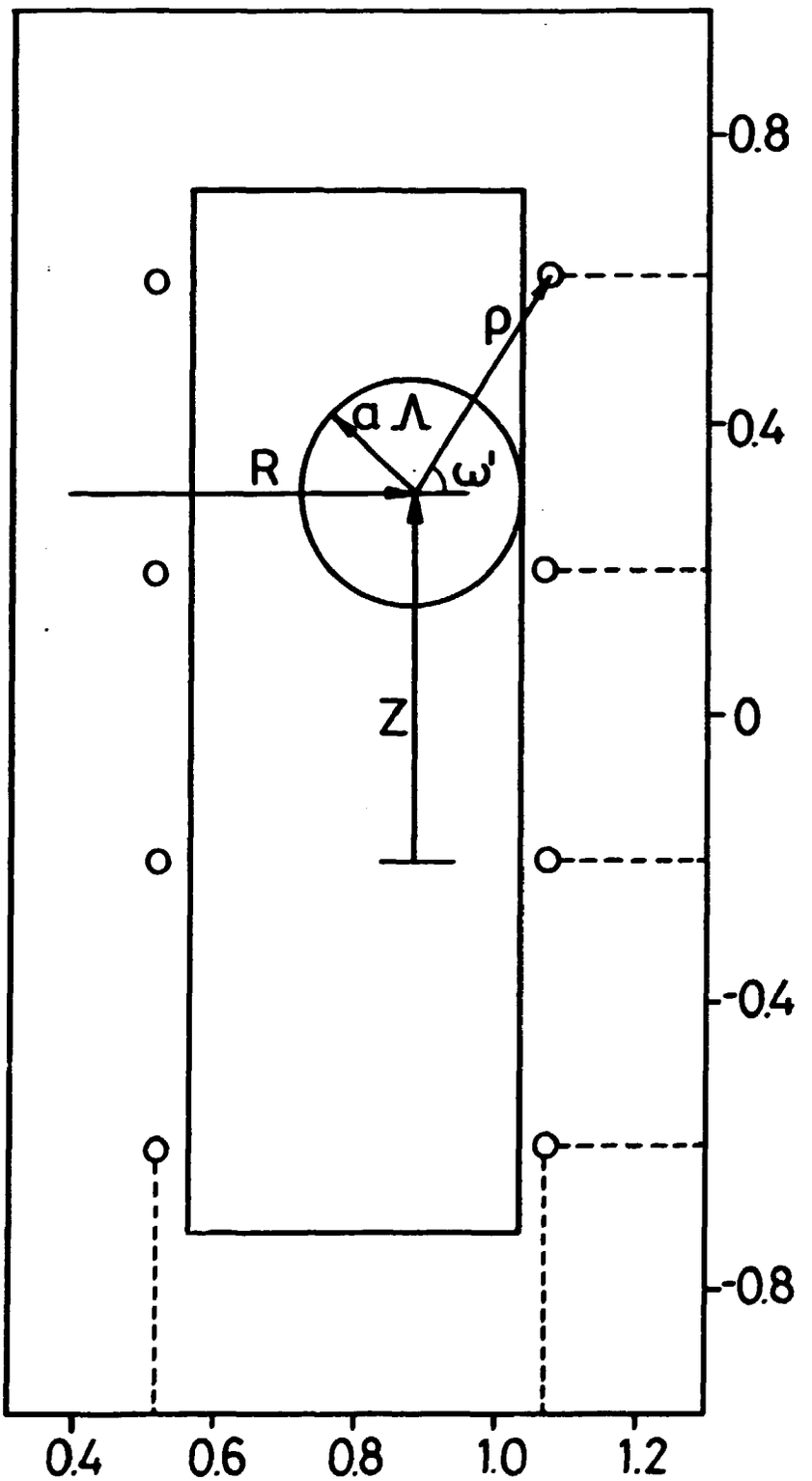

LAYER #1
INPUT
LAYER #2
HIDDEN
LAYER #3
OUTPUT
W_{11}^{23}
1
2
3
1
N_3
Σ_3
N_1
$W_{N_1N_2}^{12}$
Σ_2
N_2
BIAS
S_1
BIAS
S_2
Linear summation
Linear fan-out
Sigmoidal compression
FORWARD PASS NETWORK

Fig. 1

Fig. 2



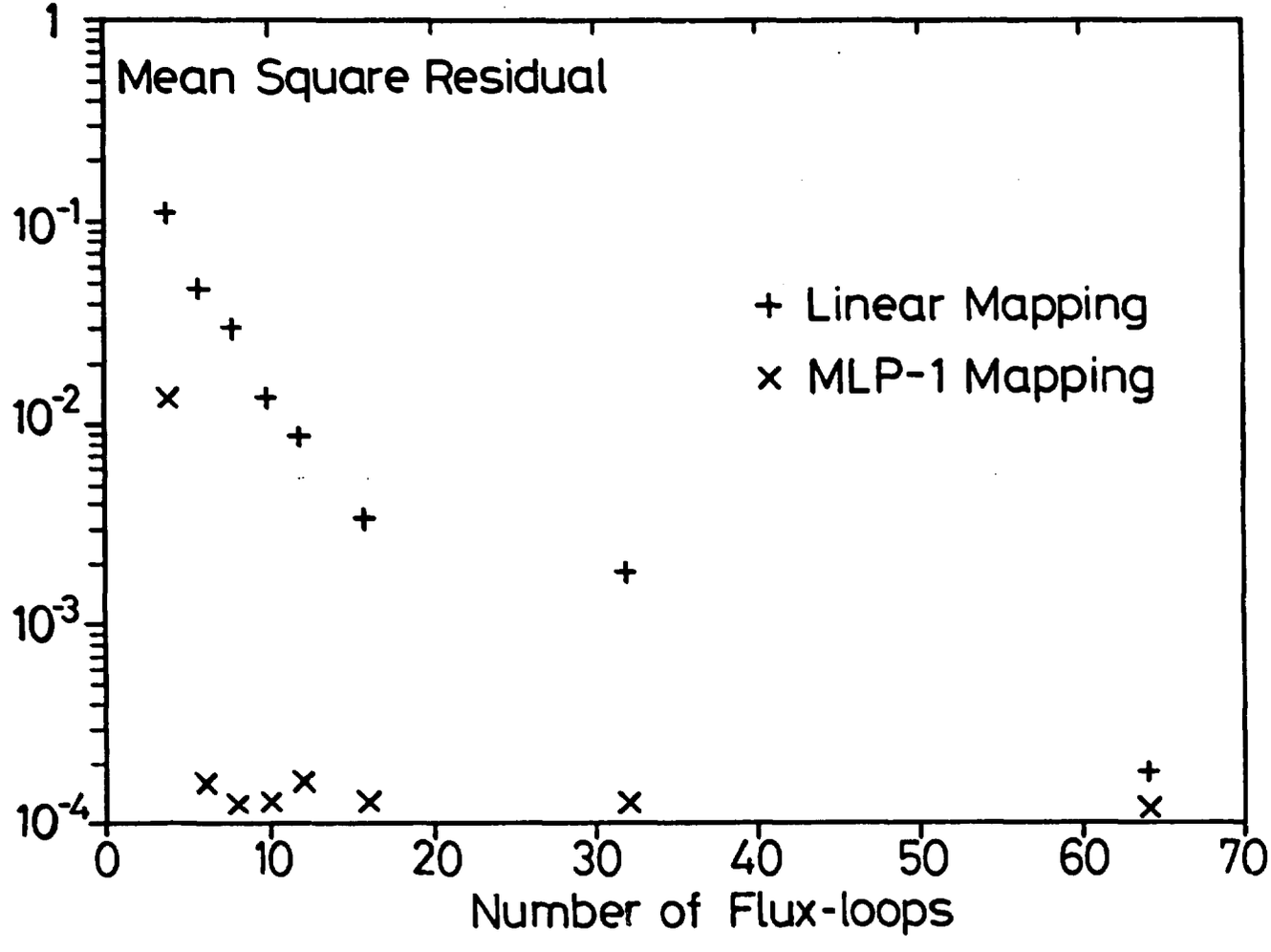

Mean Square Residual
+ Linear Mapping
× MLP-1 Mapping
1
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
10
20
30
40
50
60
70
Number of Flux-loops

Fig. 3

Fig. 4

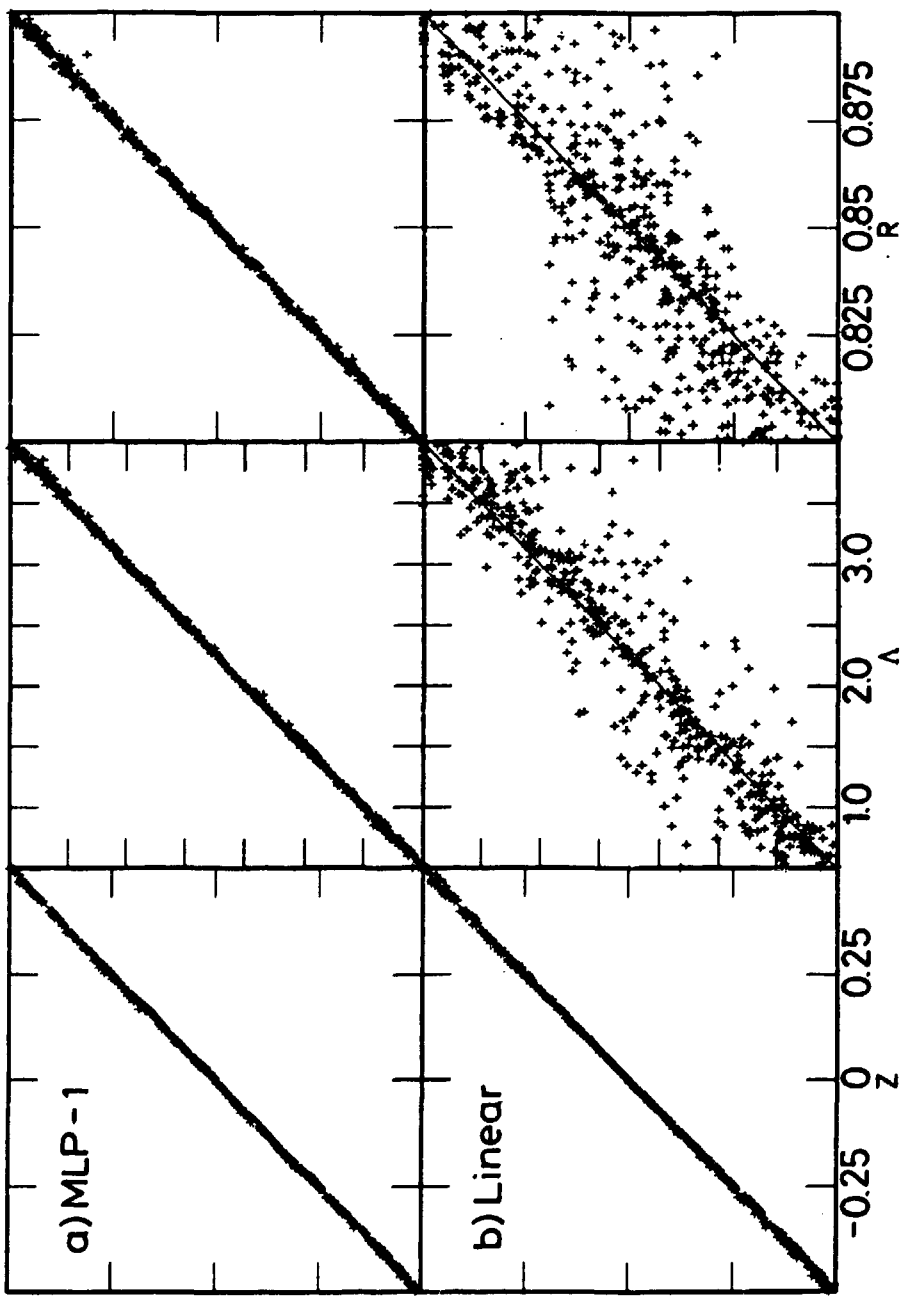
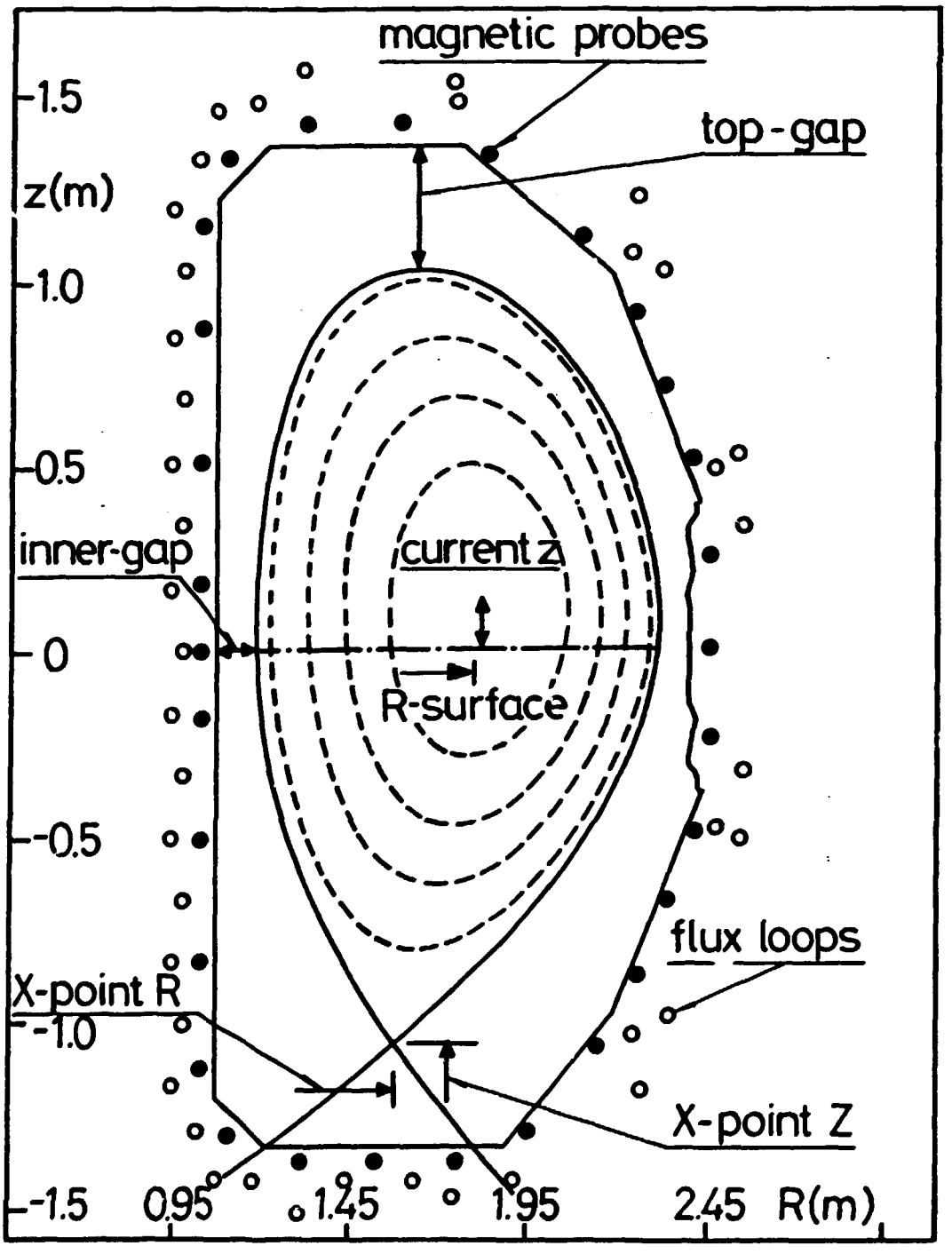


Fig. 5


magnetic probes
top-gap
1.5
z(m)
1.0
0.5
inner-gap
current z
0
R-surface
-0.5
flux loops
X-point R
-1.0
X-point Z
-1.5
0.95
1.45
1.95
2.45
R(m)

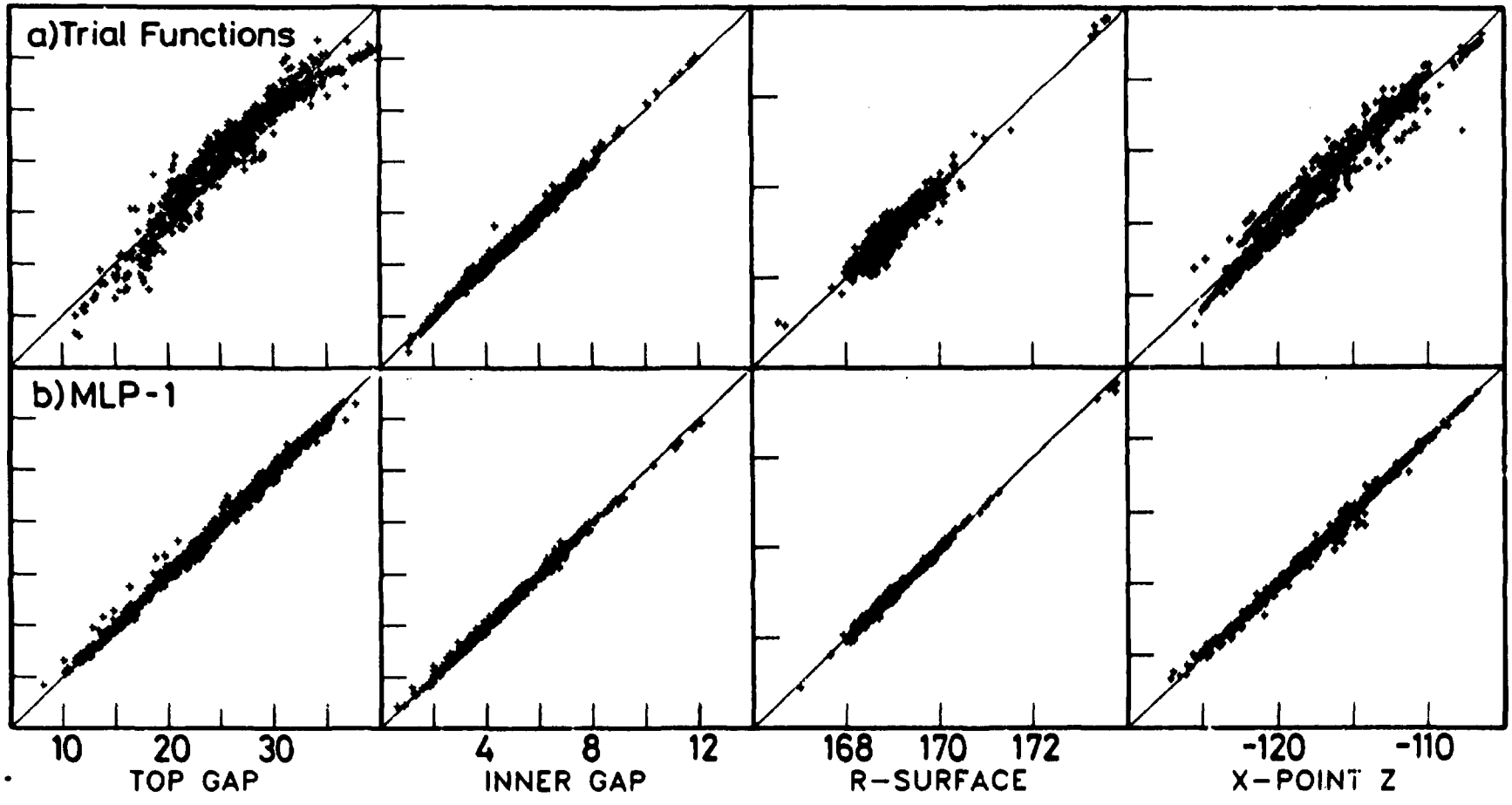

a)Trial Functions
b)MLP-1
10
20
30
TOP GAP
4
8
12
INNER GAP
168
170
172
R-SURFACE
-120
-110
X-POINT Z

Fig. 6

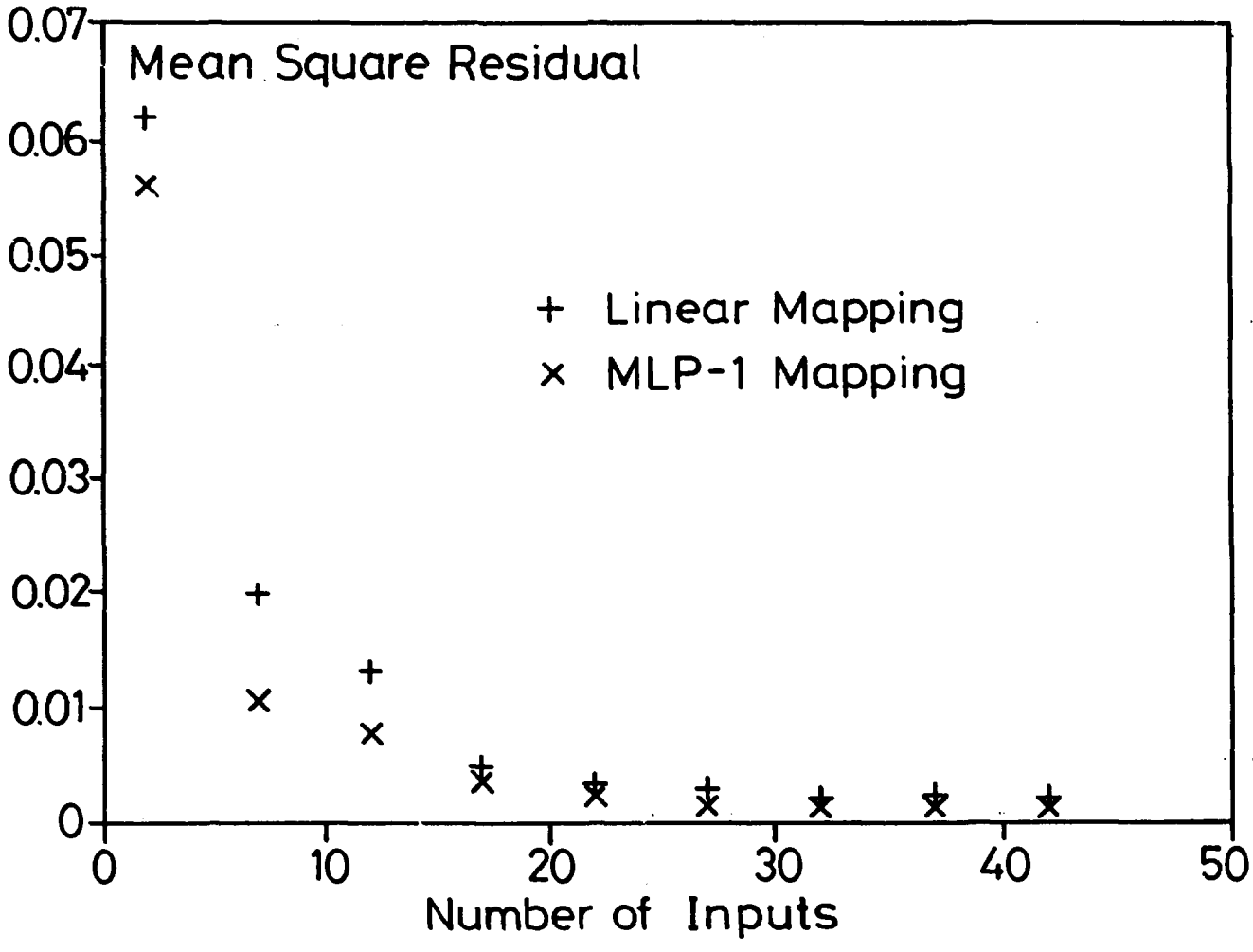

Mean Square Residual
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
+ Linear Mapping
× MLP-1 Mapping
0
10
20
30
40
50
Number of Inputs

Fig. 7