Bayesian Analysis Users Guide
Release 4.00, Manual Version 1

G. Larry Bretthorst
Biomedical MR Laboratory
Washington University School Of Medicine,
Campus Box 8227
Room 2313, East Bldg.,
4525 Scott Ave.
St. Louis MO 63110
http://bayes.wustl.edu
Email: larry@bayes.wustl.edu

October 21, 2016
8 Bayes Analyze .......................... 155
8.1 Bayes Model .................................. 159
8.2 The Bayes Analyze Model Equation .................................. 161
8.3 The Bayesian Calculations .................................. 167
8.4 Levenberg-Marquardt And Newton-Raphson .................................. 171
8.5 Outputs From The Bayes Analyze Package .................................. 176
8.5.1 The “bayes.params.nnnn” Files .................................. 177
8.5.1.1 The Bayes Analyze File Header .................................. 178
8.5.1.2 The Global Parameters .................................. 182
8.5.1.3 The Model Components .................................. 184
8.5.2 The “bayes.model.nnnn” Files .................................. 185
8.5.3 The “bayes.output.nnnn” File .................................. 186
8.5.4 The “bayes.probabilities.nnnn” File .................................. 190
8.5.5 The “bayes.log.nnnn” File .................................. 193
8.5.6 The “bayes.status.nnnn” and “bayes.accepted.nnnn” Files .................................. 196
8.5.7 The “bayes.model.nnnn” File .................................. 197
8.5.8 The “bayes.summary1.nnnn” File .................................. 198
8.5.9 The “bayes.summary2.nnnn” File .................................. 199
8.5.10 The “bayes.summary3.nnnn” File .................................. 200
8.6 Bayes Analyze Error Messages .................................. 200
9 Big Peak/Little Peak .......................... 207
9.1 The Bayesian Calculation .................................. 209
9.2 Outputs From The Big Peak/Little Peak Package .................................. 216
10 Metabolic Analysis .......................... 219
10.1 The Metabolic Model .................................. 223
10.2 The Bayesian Calculation .................................. 225
10.3 The Metabolite Models .................................. 228
10.3.1 The IPGD_D2O Metabolite .................................. 228
10.3.2 The Glutamate.2.0 Metabolite .................................. 232
10.3.3 The Glutamate.3.0 Metabolite .................................. 235
10.4 The Example Metabolite .................................. 236
10.5 Outputs From The Bayes Metabolite Package .................................. 238
11 Find Resonances .......................... 239
11.1 The Bayesian Calculations .................................. 241
11.2 Outputs From The Bayes Find Resonances Package .................................. 246
12 Diffusion Tensor Analysis .......................... 247
12.1 The Bayesian Calculation .................................. 249
12.2 Using The Package .................................. 254
13 Big Magnetization Transfer .......................... 259
13.1 The Bayesian Calculation .................................. 259
13.2 Outputs From The Big Magnetization Transfer Package .................................. 262
Outlier Detection 475

Bibliography 479
List of Figures

1.1 The Start Up Window ........................................... 23
1.2 Example Package Exponential Interface .......................... 25

2.1 Installation Kit For The Bayesian Analysis Software ................. 31

3.1 The Start Up Window ........................................... 34
3.2 The Files Menu ............................................... 35
3.3 The Files/Load Image Submenu .................................. 37
3.4 The Packages Menu ............................................ 41
3.5 The Working Directory Menu ................................... 46
3.6 The Working Directory Information Popup ........................ 47
3.7 The Settings Pull Down Menu .................................. 47
3.8 The McMC Parameters Popup ................................... 48
3.9 The Edit Server Popup .......................................... 49
3.10 The Submit Job Widgets ....................................... 51
3.11 The Server Widgets Group .................................... 52
3.12 The Ascii Data Viewer ......................................... 54
3.13 The Fid Data Viewer .......................................... 55
3.14 Fid Data Display Type ........................................ 56
3.15 Fid Data Options Menu ....................................... 58
3.16 The Image Viewer .............................................. 60
3.17 The Image Viewer Right Mouse Popup Menu ......................... 61
3.18 The Prior Probability Viewer .................................. 66
3.19 The Fid Model Viewer ......................................... 69
3.20 The Plot Results Viewer ....................................... 72
3.21 Plot Information Popup ....................................... 73
3.22 The Text Results Viewer ....................................... 75
3.23 The Bayes Condensed File .................................... 78
3.24 Data, Model, And Resid Plot .................................. 81
3.25 The Parameter Posterior Probabilities ........................... 82
3.26 The Maximum Entropy Histograms ............................... 84
3.27 The Parameter Samples Plot ................................... 85
3.28 Posterior Probability Vs Parameter Value ........................... 86
3.29 Posterior Probability Vs Parameter Value, A Skewed Example .......... 87
3.30 The Expected Value Of The Logarithm Of The Likelihood ............. 89
List of Tables

8.1 Multiplet Relative Amplitudes ........................................ 165
8.2 Bayes Analyze Models ..................................................... 181
8.3 Bayes Analyze Short Descriptions .................................... 195
Chapter 17

Unknown Polynomial Order

The Unknown Polynomial Order package fits polynomials to two column Ascii data when both the order of the polynomial and the polynomial coefficients are unknown. The interface to this package is shown in Figure 17.1. This interface differs from most others in one respect, there are no parameter ranges to enter, so use of the interface is particularly simple. To use this package, you must do the following:

Select  the Polynomial Models package from the Package menu. When selected this menu will bring up the “Given” and “Unknown” polynomial model interface.

Check  the “Unknown Order” box to select the Unknown Polynomial Order package. When this check box is activated the “Set Order” widget becomes inactive. This is illustrated in Fig. 17.1 where the “Unknown Order” has been checked, and, consequently, the “Set Order” widget has been grayed out.

Load  one two column Ascii data sets. The data may be loaded using the Files menu. You can also load an arrayed Fid and then use a single cursor to mark the center of a peak and use the “Get Peak” button on the bottom right of the Fid viewer. Finally, the “Files/Load Ascii/Bayes Analyze” button can be used to load an Ascii data set from the amplitudes estimated by Bayes Analyze. When a data set is successfully loaded the data is plotted in the Ascii Data viewer. This package does not allow you to run with multiple data sets. If you attempt to do so, you will be prompted to remove all but a single file.

Select  the server that is to process the analysis.

Check  the status of the selected server to determine if the server is busy, change to another server if the selected server is busy.

Run  the the analysis on the selected server by activating the Run button.

Get  the the results of the analysis by activating the Get Job button. If the analysis is running, this button will return the Accepted report containing the status of the current run. Otherwise, it will fetch and display the results from the current analysis.
Figure 17.1: Unknown Polynomial Order Package Interface

Figure 17.1: This panel is the interface to both the Given and Unknown Polynomial Order packages. Data has already been loaded. Note that in this example, the “Unknown Order” check box has been set. Consequently, the “Set Order” spinner has been deactivated. Because of the way this calculation is done very high orders are possible and numerically stable. However, the high orders, above 40, require very high signal-to-noise and even then roundoff errors degrade the accuracy to 4 or 5 decimal places.
17.1 Bayesian Calculations

The Unknown Polynomial Order model is just that, it’s a model in which a polynomial is fit to the data:

\[ d_i = \sum_{j=0}^{m} A_j t_i^j + n_i \]  

(17.1)

where \( A_j \) is the amplitude of the \( j \)th polynomial, \( m \) is the unknown order of the polynomial expansion, and \( n_i \) represents noise in the data. As in Chapter 16, we introduce a change of function and a change of variables and we refer the reader to that chapter for a discussion of the change of function and variables. The change of function is to orthonormal polynomials designated by \( G_j(t_i) \), so the expansion given in Eq. (17.1) becomes:

\[ d_i = \sum_{j=0}^{m} B_j G_j(t_i) + n_i \]  

(17.2)

where \( B_j \) are the amplitudes in the orthonormal expansion. Note this change of function and variables is an identity, so

\[ \sum_{j=0}^{m} A_j t_i^j = \sum_{j=0}^{m} B_j G_j(t_i). \]  

(17.3)

The Bayesian calculation is implemented using Markov chain Monte Carlo with simulated annealing to draw samples from the joint posterior probability for the parameters, \( P(mB_0B_1 \ldots B_m | D I) \). From these samples we then compute the marginal posterior probabilities for the amplitudes and the polynomial order. The joint posterior probability for the parameters is computed by application of Bayes’ theorem

\[ P(mB_0B_1 \ldots B_m | D I) \propto P(m|I) P(D|mB_0B_1 \ldots B_m | I) \]  

(17.4)

where \( P(mB_0B_1 \ldots B_m | I) \) is the joint prior probability for the amplitudes and the polynomial order, and \( P(D|mB_0B_1 \ldots B_m | I) \) is the direct probability for the data given the parameters and the polynomial order. We factor the joint prior probability for the parameters, \( P(mB_0B_1 \ldots B_m | I) \), into a series of independent prior probabilities:

\[ P(mB_0B_1 \ldots B_m | I) = P(m|I) \prod_{j=0}^{m} P(B_j|I) \]  

(17.5)

where \( P(m|I) \) is the prior probability for the polynomial order and \( P(B_j|I) \) is the prior probability for the \( j \)th amplitude. Substituting, Eq. (17.5) into Eq. (17.4) one obtains

\[ P(mB_0B_1 \ldots B_m | D I) \propto P(m|I) \left[ \prod_{j=0}^{m} P(B_j|I) \right] P(D|mB_0B_1 \ldots B_m | I) \]  

(17.6)

as the joint posterior probability for all of the parameters, including the polynomial order.
17.1.1 Assigning Priors

Before we assign the likelihood, we are going to assign the prior probability for the polynomial order, $P(m|I)$, and the prior probability for the amplitudes, the $P(B_j|I)$. The prior probability for the polynomial order was assigned as a discrete Gaussian with lower bound zero, an upper bound of 50, a mean value of 5, and a standard deviation of 10:

$$P(m|I) = \begin{cases} \frac{1}{C} \exp \left\{ -\frac{(5 - m)^2}{2 \times 10^2} \right\} & m \in \{0, 1, \ldots, 50\} \\ 0 & \text{otherwise} \end{cases},$$

with the normalization constant $C$ set so that the sum over the total models is one:

$$C = \sum_{m=0}^{50} \exp \left\{ -\frac{(5 - m)^2}{2 \times 10^2} \right\}. \quad (17.8)$$

Which expresses a belief that the polynomial order should be small, we think it very unlikely that the polynomial order would be as large as 50; but we think it reasonably possible for the order to be in the tens or twenties.

The prior probabilities for the amplitudes will be assigned exactly the same way they were when the did the Given Polynomial Order package, Chapter 16. That prior was given by Eq. 16.15 and we simply use that prior here:

$$P(B_j|I) = \left( \frac{2}{\pi \delta^2} \right)^{1/2} \exp \left\{ -\frac{(T_j - B_j)^2}{2 \delta^2} \right\} \quad (17.9)$$

where $\delta$ is the standard deviation of this prior probability and indicates how strongly we believe the expected amplitude is $T_j$. How we set $\delta$ is explained shortly. The expected amplitude, $T_j$, is given by

$$T_j \equiv \sum_{i=1}^{N} d_i G_j(t_i). \quad (17.10)$$

In Chapter 16 when this prior was used, we knew the order of the expansion polynomial and thus could determine the mean-square residual. We could use the mean-square residual to set $\delta$ to a value much wider than any amplitude supported by the data. So this prior probability just acted as a guide to the Markov chain Monte Carlo simulations. Here setting $\delta$ is harder because we don’t know which model to use. However, we still want to set $\delta$ to a value that will guide the Markov chain Monte Carlo simulations but not make $\delta$ so large that the simulations never converge. In Chapter 16 we noted that because these amplitudes appear in the model in a linear fashion, we could solve the problem analytically, we don’t have to use Markov chain Monte Carlo at all. The only reason for using a Markov chain is for consistency with the other packages in our Bayesian Analysis software. However, there is nothing to stop us from computing $P(m|DI)$ analytically and using that to set $\delta$. Without going into the details of this calculation, the posterior probability for polynomial of order $m$ is given by:

$$P(m|DI) = P(m|I) \Gamma \left( \frac{m}{2} \right) \Gamma \left( \frac{N - m}{2} \right) \left[ \frac{\bar{h}^2}{m} \right]^{-\frac{m}{2}} \left[ \frac{d^2 - \bar{h}^2}{2} \right]^{-(N-m)/2} \quad (17.11)$$
where \( P(m|I) \) is given by Eq. (17.7) and the prior probability for the amplitudes was assigned as a normalized unbounded Gaussian with mean zero and standard deviation \( \gamma \); which was marginalized out of the problem using a series of approximations given in [2] The quantity, \( \tilde{d}^2 - \tilde{h}_m^2 \), is the total-squared residual given a polynomial of order \( m \). The sufficient statistic, \( \tilde{h}_m^2 \), is the total-squared projection of the data onto the given polynomial model and is defined as

\[
\tilde{h}_m^2 \equiv \sum_{k=0}^{m} T_k^2. \tag{17.12}
\]

The expected standard deviation of the noise independent of the model order is given by:

\[
\sqrt{\langle \sigma^2 \rangle} = \max_{m=0}^{M_{\text{max}}} P(m|DI) \sqrt{\frac{\tilde{d}^2 - \tilde{h}_m^2}{N}}. \tag{17.13}
\]

Finally, \( \delta \) was set to

\[
\delta = 10 \sqrt{\langle \sigma^2 \rangle}. \tag{17.14}
\]

While rather complicated, this calculation was used for two reasons: I needed an estimate of the standard deviation of the noise, which this gives by simple straightforward calculation; and I needed a way to determine where the maximum of the posterior probability for the polynomial order was. I needed this maximum to determine where to center the distribution of simulations that is printed out while this program is running. The problem is illustrated in Fig. 17.2. The maximum order of the polynomial is 50, but there is only room to print out 10 of these probabilities. So the output window must be shifted to cover the maximum posterior probability for the polynomial order. To do that, I needed to know where the maximum was. This calculation solved both of these problems at one time and it did so using Bayesian probability theory. See my book, [2], for more on this calculation and where each of these terms comes from.

### 17.1.2 Assigning The Joint Posterior Probability

Having assigned the prior probabilities, we can now proceed with assigning the joint posterior probability for the parameters, Eq. (17.6). First, however, we assign the direct probability for the data. The direct probability, \( P(D|mB_0B_1 \ldots B_mI) \), is a marginal probability and is computed from the joint probability for the data and the standard deviation of the noise

\[
P(D|mB_0B_1 \ldots B_mI) = \int P(\sigma D|mB_0B_1 \ldots B_mI)d\sigma \tag{17.15}
\]

which we factor as

\[
P(D|mB_0B_1 \ldots B_mI) = \int P(\sigma|I)P(D|\sigma mB_0B_1 \ldots B_mI)d\sigma. \tag{17.16}
\]

Assigning a Jeffreys’ prior to \( P(\sigma|I) \) and a Gaussian likelihood, one obtains

\[
P(mB_0B_1 \ldots B_m|DI) \propto P(m|I) \left[ \prod_{j=0}^{M} P(B_j|I) \right] \left[ \int \frac{1}{\sigma} (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left\{ -\frac{Q}{2\sigma^2} \right\} d\sigma \right] \tag{17.17}
\]
The table below shows the distribution of models on the console log for two different phases: Annealing and Sampling.

**Annealing**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Annl Param</th>
<th>&lt;Likelihood&gt;</th>
<th>&lt;StdDevLike&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>-2.8665E+01</td>
<td>-1.4089E+02</td>
<td>4 5 6 7 8 9 10 11 12</td>
</tr>
<tr>
<td>0.004</td>
<td>-3.1841E+01</td>
<td>-5.2177E+01</td>
<td>5 5 7 4 3 5 2 4 0 0</td>
</tr>
<tr>
<td>0.007</td>
<td>-3.1615E+01</td>
<td>9.4579E+00</td>
<td>6 5 14 4 5 2 3 1 1 0</td>
</tr>
<tr>
<td>0.012</td>
<td>-3.2178E+01</td>
<td>5.4344E+01</td>
<td>4 7 15 8 4 4 0 1 1 0</td>
</tr>
<tr>
<td>0.017</td>
<td>-3.4670E+01</td>
<td>1.0615E+02</td>
<td>2 6 14 13 7 3 0 1 1 0</td>
</tr>
<tr>
<td>0.025</td>
<td>-3.6671E+01</td>
<td>1.3447E+02</td>
<td>0 4 19 8 7 6 3 1 0 0</td>
</tr>
<tr>
<td>0.039</td>
<td>-3.6100E+01</td>
<td>1.5631E+02</td>
<td>0 0 21 13 7 2 3 2 0 0</td>
</tr>
<tr>
<td>0.065</td>
<td>-3.4743E+01</td>
<td>1.7613E+02</td>
<td>0 0 16 18 7 6 1 0 0 0</td>
</tr>
<tr>
<td>0.088</td>
<td>-3.3993E+01</td>
<td>1.9458E+02</td>
<td>0 0 11 22 12 3 0 0 0 0</td>
</tr>
<tr>
<td>0.109</td>
<td>-3.4516E+01</td>
<td>2.0296E+02</td>
<td>0 0 8 24 10 4 2 0 0 0</td>
</tr>
</tbody>
</table>

**Sampling**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Frac</th>
<th>&lt;Likelihood&gt;</th>
<th>&lt;StdDevLike&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.500</td>
<td>0.</td>
<td>-3.1813E+01</td>
<td>2.6462E+02</td>
</tr>
<tr>
<td>0.550</td>
<td>0.</td>
<td>-3.1991E+01</td>
<td>2.6438E+02</td>
</tr>
<tr>
<td>0.600</td>
<td>0.</td>
<td>-3.1907E+01</td>
<td>2.6411E+02</td>
</tr>
<tr>
<td>0.650</td>
<td>0.</td>
<td>-3.1989E+01</td>
<td>2.6459E+02</td>
</tr>
<tr>
<td>0.700</td>
<td>0.</td>
<td>-3.1817E+01</td>
<td>2.6430E+02</td>
</tr>
<tr>
<td>0.750</td>
<td>0.</td>
<td>-3.1904E+01</td>
<td>2.6432E+02</td>
</tr>
<tr>
<td>0.800</td>
<td>0.</td>
<td>-3.1994E+01</td>
<td>2.6414E+02</td>
</tr>
<tr>
<td>0.850</td>
<td>0.</td>
<td>-3.1992E+01</td>
<td>2.6429E+02</td>
</tr>
<tr>
<td>0.900</td>
<td>0.</td>
<td>-3.1909E+01</td>
<td>2.6386E+02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.</td>
<td>-3.1903E+01</td>
<td>2.6447E+02</td>
</tr>
</tbody>
</table>

**Sampling**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Frac</th>
<th>&lt;Likelihood&gt;</th>
<th>&lt;StdDevLike&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.500</td>
<td>0.</td>
<td>-3.1813E+01</td>
<td>2.6462E+02</td>
</tr>
<tr>
<td>0.550</td>
<td>0.</td>
<td>-3.1991E+01</td>
<td>2.6438E+02</td>
</tr>
<tr>
<td>0.600</td>
<td>0.</td>
<td>-3.1907E+01</td>
<td>2.6411E+02</td>
</tr>
<tr>
<td>0.650</td>
<td>0.</td>
<td>-3.1989E+01</td>
<td>2.6459E+02</td>
</tr>
<tr>
<td>0.700</td>
<td>0.</td>
<td>-3.1817E+01</td>
<td>2.6430E+02</td>
</tr>
<tr>
<td>0.750</td>
<td>0.</td>
<td>-3.1904E+01</td>
<td>2.6432E+02</td>
</tr>
<tr>
<td>0.800</td>
<td>0.</td>
<td>-3.1994E+01</td>
<td>2.6414E+02</td>
</tr>
<tr>
<td>0.850</td>
<td>0.</td>
<td>-3.1992E+01</td>
<td>2.6429E+02</td>
</tr>
<tr>
<td>0.900</td>
<td>0.</td>
<td>-3.1909E+01</td>
<td>2.6386E+02</td>
</tr>
<tr>
<td>0.950</td>
<td>0.</td>
<td>-3.1903E+01</td>
<td>2.6447E+02</td>
</tr>
</tbody>
</table>

Figure 17.2: While the unknown Polynomial Order package is running, it prints a listing that shows the distribution of the model indicator as a function of the annealing parameter. The 10 columns to the right are the number of simulations in model 0, 1, etc. As the annealing parameter increases these should cluster into one or two columns and the distribution of these simulations is the posterior probability for the polynomial order. In this example the data were a 6th order polynomial. Notice that as soon as the annealing parameter begins to increase the simulations quickly move to high order models and eventually they usually all end up in the 6th order polynomial model.
where we have left the prior probabilities in their symbolic form. Evaluating the integral over $\sigma$ one obtains

$$P(m B_0 B_1 \ldots B_m | D I) \propto P(m | I) \prod_{j=0}^{m} P(B_j | I) \left[ \frac{Q}{2} \right]^{-\frac{N}{2}}$$

(17.18)

as the posterior probability for the parameters including the polynomial order, where

$$Q \equiv \sum_{i=1}^{N} \left( d_i - \sum_{j=0}^{m} B_j G_j(t_i) \right)^2$$

(17.19)

In evaluating the integral over $\sigma$ there were a number of constants that were dropped. In model selection problems that is usually a bad thing to do and will, almost always, cause problem. Here we could do it because each polynomial model contains exactly the same constants and so they always cancel. Finally, substituting the prior probability for the polynomial order, Eq. (17.7), the prior probability for the amplitudes, Eq. (17.9) into Eq. (17.18), the joint posterior probability for the parameters is given by:

$$P(m B_0 B_1 \ldots B_m | D I) \propto \exp \left\{ -\frac{(5 - m)^2}{2 \times 10^2} \right\} \left[ \prod_{j=0}^{m} (2\pi \delta^2)^{-\frac{1}{2}} \exp \left\{ -\frac{(T_j - B_j)^2}{2\delta^2} \right\} \right] \left[ \frac{Q}{2} \right]^{-\frac{N}{2}}.$$

(17.20)

It is this joint probability density function that is targeted by the Markov chain Monte Carlo simulations.

If one were to compute the posterior probability for the polynomial order using Eq. (17.20) and compare it to that given by Eq. (17.11). You would find you get different results. That's because in computing these two sets of equations we used slightly different prior probabilities for the amplitudes. While these prior probabilities were not much different, they are nonetheless different and that difference would manifest itself as a slight difference in the final calculations. As far as which is right, they are both correct given the two sets of prior information. Regardless, they won't differ by much and almost certainly, given the discrete nature of the posterior probability, won't differ at all after you normalize the final posterior probability.

### 17.2 Outputs From the Unknown Polynomial Order Package

The Text outputs from the Unknown Polynomial Order package consist of: “Bayes.prob.model,” “BayesPolUnknown.mcmc.values,” “Bayes.params,” “Console.log” see Fig. 17.2, “Bayes.accepted” and a condensed output file “Bayes.Condensed.File.” These output files can be viewed using the Text Viewer or they can be viewed using File Viewer by navigating to the current working directory and then selecting the files. The format of the mcmc.values report is discussed in Appendix D and the other reports are discussed in Chapter 3.

The main new output from the Unknown Polynomial Order package is a plot of the posterior probability for the polynomial order, Fig. 17.3. The data used to generate this figure are the
Figure 17.3: The main new output in the Unknown Polynomial Order Package is the posterior probability for the polynomial order, here called polynomial number. This output figure contains 10 probabilities centered around the peak in the posterior probability.
Polynomials.6th.order.dat distributed in the Bayes.test.data. This figure consists of a bar chart of the posterior probability. However, there are only 10 output probabilities, while the posterior probability contains a maximum of 51 probabilities (orders 0 through 50). So these probabilities are centered around the location of the maximum, see Section 17.1.1 for a discussion of how this maximum is located. Usually all of the probability is concentrated in one or two probabilities around the maximum with an abrupt lower bound and a more gentle drop off as you go to higher orders. Consequently, after locating the maximum of the posterior probability for the model order, the lowest output probability is set 3 orders below the maximum and the maximum output probability is 6 orders above the maximum probability order. All this is illustrated in Fig. 17.3, there the maximum is 6, the lowest output order is \(6 - 3 = 3\), and the highest is \(6 + 6 = 12\).
Bibliography


[45] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller (1953), “Equation of State Calculations by Fast Computing Machines,” Journal of Chemical Physics. The previous link is to the American Institute of Physics and if you do not have access to Science Sitations you many not be able to retrieve this paper.


