Analog And Digital Signal Processing, Calibration And Image Processing In Radio Telescopes

Rohan Chabukswar

Department of Physics
Indian Institute of Technology Bombay

November 29, 2006

Outline

Outline (cont.)

Birth Of Radio Astronomy

- Radio astronomy began in 1932 with Karl Jansky's discovery of a strong source of radio static originating from the central region of the Milky Way.
- A young engineer in Wheaton, IL, USA named Grote Reber constructed a fully steerable parabolic dish antenna and was able to construct a map of the sky covered by his antenna.
- After World War II, when many scientists found themselves free of military concerns and armed with a wealth of new technology which sprung from the conflict, especially Radar technology. They used this knowledge to pursue the new field of Radio Astronomy.

Historically Important Instruments

- A radio analog of the Michelson Interferometer was constructed by Ryle and Vonberg in 1964.
- The sea interferometer was used in Australia with already installed horizon-looking radar antennas. Radiation from sources rising over the eastern horizon was received reflected from the sea and directly, causing interference.
- During the 1950s and early 1960s most radio telescopes were of the non-tracking array type, where the rotation of the earth is used scan the source.
- These were succeeded by one dimensional tracking, and then two dimensional arrays. In 1976, a new technique of interferometry called Very Long Baseline Interferometry (VLBI) was developed.

Radio Astronomy Instruments in India

- The Ooty Radio Telescope in Udhagamandalam, India is of the non-tracking type, with an array of more than 1000 dipoles approximately 0.5 kilometer long in the North-South direction.It was constructed in 1960.
- The largest two dimensional tracking array in the meter-wave range, is the Giant Meter-wave Radio Telescope, situated at Khodad near Pune, India.
- Fourteen of the thirty dishes in GMRT are located more or less randomly in a compact central array in a region of about 1 square kilometer. The remaining sixteen dishes are spread out along the 3 arms of an approximately ‘Y’-shaped configuration over a much larger region, with the longest interferometric baseline of about 25 km.

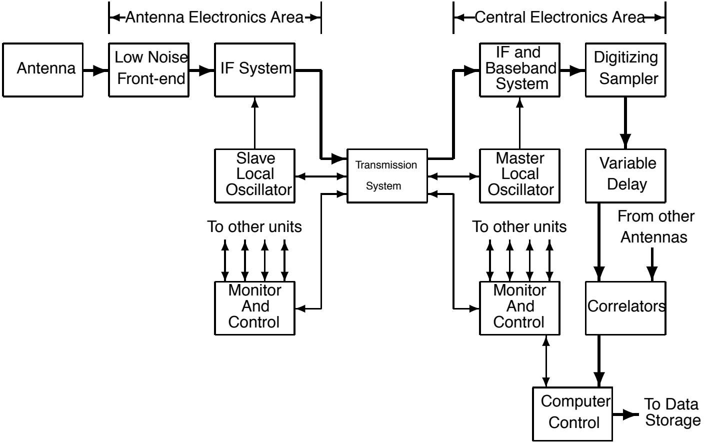
Technical Terms In Radio Astronomy

- Since the field of radio astronomy was started by electronics engineers, the terms used in radio astronomy are taken from engineering rather than astronomy.
- The voltages induced by cosmic-source radiation are called signals. They are generated by natural processes and have the form of Gaussian random noise.
- These characteristics are mostly stationary with time, and the signals are ergodic.
- The noise/signal power in a system is specified in terms of the temperature of a resistor load that would produce an equal power level in a noise free receiver.

Technical Terms In Radio Astronomy

- The flux of signals coming from the source as a function of $(\zeta, \ \eta)$ which are coordinates on celestial sphere, is called the brightness distribution of the region.
- The flux as a function of spatial frequency in wavelength units or $(u, \ v)$, is called visibility.
- The signals from different antennas, when multiplied together with a specific delay (called cross-correlation) gives us the visibilities.
- The brightness distribution can be generated by just a Fourier Transform of the visibilities.

The Receiving System Of A Typical Synthesis Array



Principle Subsystems

Low Noise Input Stages

- To achieve high sensitivity in a receiving system, the noise in the antenna and the input stages of the receiver must be as low as possible.

$$T_S = T'_A + (L_1 - 1)T + T_{R1}L_1 + T_{R2}L_1G_1^{-1} + T_{R2}L_1(G_1G_2)^{-1} + \ldots$$

- Modern methods to minimize noise temperature involve cryogenic cooling of amplifiers and mixer stages from input to the point where noise from succeeding stages is unimportant.

Principle Subsystems

Local Oscillator

- Frequencies of the oscillators at different antennas must be maintained in phase to preserve coherence.
- Mixing enables major part of signal processing to be performed at an intermediate frequency appropriate for amplification, transmission, filtering, delaying, recording etc. A mixer adds/subtracts the local oscillator frequency to the signal frequency.
- To control frequency of sinusoidal fringe variations in correlator output, a continuous phase change can be inserted into one of local oscillator signals, called fringe rotation. These phase shifts are also required at mixers, which are implemented by digital synthesis methods.

Principle Subsystems

IF/Signal Transmission

- This transmission is usually effected by coaxial or parallel-wire lines, a waveguide, optical fibers or a microwave radio link.
- Transmission frequencies range from 10 – 100 MHz in coaxial cables, to 10 GHz in waveguide or radio and to optical frequencies in fibers.
- Transmission system may also be used to distribute reference frequencies for local oscillator subsystems.

Principle Subsystems

Delay And Correlators

- An analog delay system contains series of binary valued switchable delay units, so that all delays from 1 to $2^n - 1$ times the smallest delay unit can be achieved.
- Analog multiplying circuits basically take the logarithms of two signals, add them, and take the antilogarithm of the sum.
- Now-a-days, digital circuitry operating at frequencies upward of 100 MHz led to the practice of digitizing final IF signal, so that delay and correlation can be done digitally with greater precision.

Phase Stability

- The phases of the local oscillators are affected by temperature variation (diurnal and annual), rotating joints, flexible cables, etc.
- Tuned filters used for selecting local oscillator frequencies are also a source of temperature-related phase variations.
- Path length variations can be monitored by phase of a signal that traverses the path and back, called round-trip phase measurement systems like Swarup And Yang system, frequency offset round-trip system, automatically correcting systems and phase locked loops with reference frequencies.

Optimum Response

- Signals in a synthesis array pass through a large number of amplifiers, filters, mixers and transmission lines, the characteristics of these instruments are impressed upon the signals.
- To examine the tolerable deviations of the responses, a factor D can be defined as the signal-to-noise ratio relative to that of identical rectangular responses

$$\mathscr{D} = \frac{\int_0^\infty H_m(\nu)\, H_n^*(\nu)\, d\nu}{\sqrt{\Delta\nu \int_0^\infty |H_m(\nu)|^2\, |H_n(\nu)|^2\, d\nu}}.$$

Tolerances On Frequency Response Variations

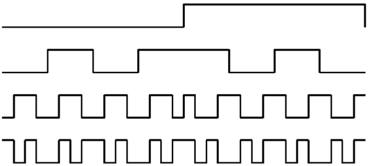
Type Of Variation	Criterion: 2.5% Degradation In SNR	Criterion: 1% Maximum Gain Error
Amplitude Slope	3 dB edge-to-edge	2.7 dB edge-to-edge
Sinusoidal Ripple	2.9 dB peak-to-peak	2.0 dB peak-to-peak
Center-Frequency Displacement	$0.05\Delta\nu$	$0.007\Delta\nu$
Phase Variation	12.8° RMS	9.1° RMS
Delay-Setting Error	$0.12/\Delta\nu$	$0.05/\Delta\nu$

Implementation Of Tolerances

- To implement all the tolerances derived above, the specifications of filters used should provide for required matching of responses and temperature effects.
- The frequency selectivity of the elements in earlier stages can then be held to the minimum required for rejecting interference, thus minimizing the effect on overall response.

Phase Switching

Walsh Functions



- An early method of obtaining the product of two voltages was to periodically reverse the phase of one. This method also reduces spurious signals by two or more orders of magnitude.
- For a number of antennas, Walsh functions are used for phase switching. The figure shows a few Walsh Functions.

Digital Signal Processing

- Accuracy of delay depends upon the accuracy of timing pulses, which is achieved more easily in digital than analog delay lines.
- There is no distortion of the signal by digital units, while it is difficult to keep the shape of frequency responses within the tolerance in analog systems.
- Multichannel output is obtained more readily in digital systems, whereas it requires filter banks in analog systems.
- Digital circuits require less on-board adjustment and are better suited for replication.

Nyquist Sampling Theorem

- For baseband with upper cutoff frequency $\Delta\nu$, the function can be fully specified by samples with sampling frequency of $2\Delta\nu$, called Nyquist rate, or greater.
- This can also be applied to bandpass functions, that is, if the spectrum is nonzero within $n\Delta\nu$ to $(n+1)\,\Delta\nu$, the Nyquist rate is again $2\Delta\nu$.
- Sampling at frequencies greater or less than Nyquist frequency is referred to as oversampling or under-sampling.

Sampled Unquantized Waveforms

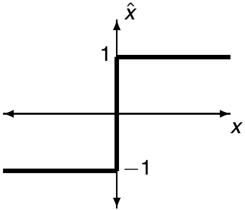
- Constructing a sampler which does not quantize the signals is not practical, but it has been studied for the sake of comparison to quantized versions.
- It can be shown that the signal-to-noise ratio with unquantized sampling is given by

$$\mathscr{R}_{sn\infty} = \frac{\rho\sqrt{\beta N_q}}{\sqrt{1 + 2\sum_{q=1}^{\infty} R_{\infty}^2 (q\tau_s)}},$$

where,

$$R_{\infty}(q\tau_s) = \frac{\beta \sin(\pi q/\beta)}{\pi q}.$$

Two Level Quantization



- A two level quantization can be achieved by simply amplifying and clipping the signal.

Two Level Quantization

- The signal-to-noise ratio is given by

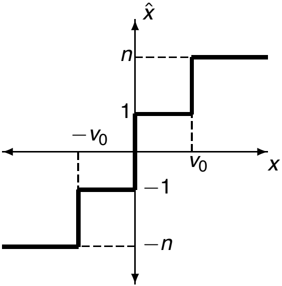
$$\mathscr{R}_{sn2} = \frac{\rho\sqrt{N}}{\pi\sqrt{1 + 2\sum_{q=1}^{\infty} R_2^2\left(q\tau_s\right)}},$$

where,

$$R = \frac{2}{\pi}\sin^{-1}\left[\frac{\beta\sin\left(\pi q/\beta\right)}{\pi q}\right].$$

- For $\beta = 1$, $\sum_{q=1}^{\infty} R_2^2\left(q\tau_s\right) = 0$, and snr is 64% of that for unquantized sampling.
- For oversampling at $\beta = 2$ and $\beta = 3$, signal-to-noise ratio is enhanced from $\mathscr{R}_{sn2}$ by a factor of 1.17 and 1.21 respectively.

Four Level Quantization



- A four-level quantization with quantization states are $-n$, -1, 1, n, threshold values $-v_0$, 0, v_0.

Four Level Quantization

- The signal-to-noise ratio is given by

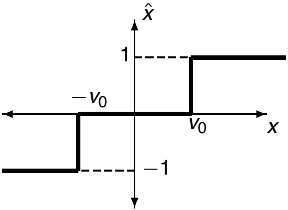
$$\mathscr{R}_{sn4} = \frac{2\rho\left[(n-1)\exp\left(-v_0^2/2\sigma^2\right)+1\right]^2\sqrt{N}}{\pi\left[\Phi+n^2(1-\Phi)\right]\sqrt{1+\sum_{q=1}^{\infty}R_4^2(q\tau_s)}},$$

$$R_4 = \frac{2\left[(n-1)\exp\left(-v_0^2/2\sigma^2\right)+1\right]^2}{\pi\left[\Phi+n^2(1-\Phi)\right]}\frac{\beta\sin(\pi q/\beta)}{\pi q},$$

$$\Phi = \mathrm{erf}\left(\frac{v_0}{\sigma\sqrt{2}}\right).$$

- Cooper (1970) has shown that $n = 3$, $v_0 = \sigma$ and $n = 4$, $v_0 = 0.95\sigma$ both result in the maximum snr of 0.88 of that for unquantized.

Three Level Quantization



- If the lower products ± 1 are counted as 0 in four level quantization, a three level quantization is obtained.

Three Level Quantization

- The signal-to-noise ratio is given by

$$\mathscr{R}_{sn3} = \frac{2\rho\sqrt{N\beta}E^2}{\pi\left(1-\Phi\right)\sqrt{1+\sum_{q=1}^{\infty} R_3^2\left(q\tau_s\right)}}.$$

- In the three bit quantization, optimum v_0 is given by 0.612σ.
- For $\beta = 1$, signal-to-noise ratio comes out to be 0.81 times as much as from unquantized case.
- For $\beta = 2$, it increases to 0.89.

Comparison Of Quantization Levels

Number Of Quantization Levels (Q)	Sensitivity Relative To Unquantized $\beta = 1$	$\beta = 2$
2	0.64	0.74
3	0.81	0.89
4	0.88	0.94

Table: Efficiency Factors For Various Quantization Schemes

- Factors that make lower quantizations preferable are complexity of further circuits and the storage space.
- 16 level quantization results in sensitivity of 97% as much as unquantized.

Digital Circuits

Digital Delay Circuits

- A digital delay circuit has a series of shift registers each with delays of 1, 2, 4, 8, etc. times the clock cycle Another way would be to use two shift registers to obtain a delay that is in variable increments of clock interval.
- RAMs and serial-to-parallel conversion can also be used.
- Finer delays can be obtained digitally by varying the timing of the sample pulse in number of steps, or by using small analog delay lines.

Digital Circuits

Quadrature Phase Shift

- Complex correlators for digital signals can be implemented by introducing quadrature phase shift in the analog signal and then using separate samplers for the actual signal and the phase shifted version.
- The Hilbert transformation that represents the phase shift can also be performed directly on the digital signal, thus eliminating all quadrature networks.
- However, loss of information must be faced if the complexity of the correlators ahead is not to be increased.

Digital Circuits

Digital Correlators

1. Correlators with both signals 2 - 3 level quantized are constructed with digital multiplier circuits.
2. Correlators with one quantization at 2 - 3 bits and one higher-quantization can be implemented by a register accumulating (adding, subtracting or ignoring based on first signal) the second signal.
3. Correlators with both numbers quantized at higher levels are implemented by having a ROM store the possible values of product, and the inputs specifying the address.

Calibration Of Visibility

- Antenna position coordinates, pointing corrections resulting from axis aligning tolerances and zero point settings of instrumental delays can be calibrated in advance.
- Effects that vary during an observation include
  1. constant component of atmospheric attenuation
  2. variation of antenna gain with altitude
  3. shadowing of one antenna by another
  4. variation of gain from ALC action
  5. phase variations in LO system
  6. variable component of atmospheric delay

Calibration Of Visibility

Use Of Calibration Sources

- The calibrator to be used should be strong, so that a good signal-to-noise ratio is obtained in short time.
- The calibrator should also be unresolved so that precise measurements of its visibility is not necessary.
- Just enough antenna spacings need to be calibrated so that all antennas are included.
- Position of calibrator should be close to that of mapped source, so that effects of atmosphere or antennas varying with pointing angle can be minimized, as also the time lost between changing antenna positions.

Calibration Of Visibility

Calibration Of Spectral Line Data

- Channel to channel differences are relatively stable with time.
- Continuum methods applied equally to all channels.
- Bandpass calibration observation is done to determine relative gains of all channels.
- The calibrator should be unresolved, provide a good signal-to-noise ratio in the channels and have flat spectrum, but need not be close to mapped source and can be observed before or after the mapping observations.

Derivation Of Brightness From Visibility

Direct And Discrete Fourier Transform

- This is the most straightforward method of obtaining brightness distribution from visibility.
- For small celestial areas, the measured visibility is the Fourier transform of the brightness distribution of the source.
- Discrete Fourier transformation can be implemented by the Fast Fourier Transform (FFT) algorithm, which reduces the time complexity of Fourier transform, but it is necessary to evaluate visibility at points in rectangular grid, with the possibility of aliasing.

Possible Error Causes

- The distribution of visibility corresponding to a suspicious feature in a map may indicate a problem related to particular group of antennas for specific time.
- Other mapping errors result from interference, correlator offset errors, etc. as also response to radio emission of the sun from the side-lobes.
- Multiplicative errors include gain constants of antennas which can arise from calibration or phase errors introduced by the atmosphere.

Visibility At Low Spatial Frequencies

- A problem common to all synthesis arrays is that the minimum antenna spacing cannot be less than diameter of antenna, and practically much greater, causing a central hole in $(u, \ v)$ coverage and broad negative side-lobes.
- The situation can be improved by inserting visibility at origin from independent measurement of flux and applying sufficient weight to it.
- Data from a smaller interferometer may also be added at the center.

Necessity

- The limited distribution of spatial frequencies and errors in measurements are the deficiencies in visibility data that limit accuracy of synthesis maps.
- Certain obvious errors like negative brightness are also introduced in the map, so it is possible to improve the result by injecting some a priori information about the source.

CLEAN Algorithm

The Algorithm

- Analytical deconvolution is not an option here, as the weighted transfer function contains a lot of zeros.
- An improved procedure must place non-zero visibility values. As these can be filled with infinitely many values, there exist infinite number of solutions.
- Judgments can be made as to what brightness distribution is reasonable and incorporated into the map, for example, extensive sinusoidal structure and negative brightness values are very obviously instrument artifacts.
- One of the more successful of such procedures is the CLEAN algorithm developed by Högbom in 1974.

CLEAN Algorithm

The Algorithm

1. Compute map and response to a point source by Fourier transformation of visibility and weighted transfer function, to get dirty map and dirty beam.
2. From highest brightness point on map subtract dirty beam with a peak amplitude of γ times the peak amplitude in map.
3. Repeat Step 2 till all significant source structure is removed from map.
4. Add removed components, in the form of clean beam responses, to the residual brightness distribution to obtain map.

CLEAN Algorithm

Implementation And Performance

- Several arbitrary choices influence CLEAN — γ, window area and the criterion for termination. γ lies between 0.1 and 0.5, and gives better results when it is at the lower range, but at a cost of higher computation time.
- A well known problem with CLEAN is the generation of spurious spots or ridges as modulation on broad features.
- A modification of CLEAN by Cornwell (1983) minimizes $\sum w_i \left|\mathscr{V}_i - \mathscr{V}_i'\right|^2 - \kappa s$ instead of $\sum w_i \left|\mathscr{V}_i - \mathscr{V}_i'\right|^2$, where κ is adjustable and s is measure of smoothness, like negative mean squared brightness of the map.

CLEAN Algorithm

Performance Of CLEAN Algorithm

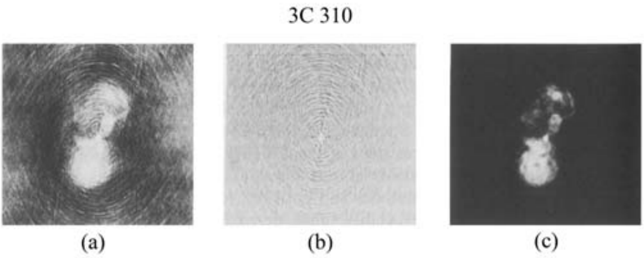


Figure: Images of the radio galaxy 3C10 from observations with the VLA.

CLEAN Algorithm
Constrained Optimization

- Constrained Optimization Techniques include a class of algorithms which produce a map constrained by maximizing some measurement of image quality.
- Maximum Entropy Method is one of these where an entropy function is defined and maximized within the constraint that Fourier transform of B should match visibility values.
- Some functions such that $\frac{d^2F}{dB'^2} < 0$ and $\frac{d^3F}{dB'^3} > 0$ were studied by Narayan and Nityananda (1982), and many of them are effective as entropy.

Mapping With Incomplete Data

Mapping With Visibility Modulus

- Circumstances in which phase data is completely missing include measurements with uncalibrated phase data, or with two antennas.
- The question of whether it is possible to obtain a unique solution for B from its autocorrelation function has been a long standing one in image processing.
- Apart from a 180° rotational ambiguity, it is indeed possible in two dimensions to obtain maps from visibility data alone, using constraints on positivity and confinement of image.

Mapping With Incomplete Data

Mapping With Uncalibrated Phase Data

- Mapping with uncalibrated phase data is of great practical interest, as instrumental limits sometimes inhibit calibration of phase and may even complicate amplitude calibration.
- Relative values of uncalibrated visibility measurements can be extracted from closure relationships between phases of three different antennas.
- Amplitude closure equations require 4 antennas each.
- Two techniques by are used, Readhead iterative mapping and self calibration.

Mapping With Incomplete Data

Performance Of Maximum Entropy Method And Self-Calibration

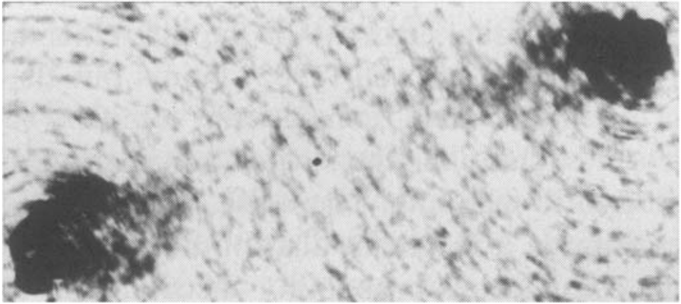


Figure: Three stages in the reduction of the observation of Cygnus A.

Mapping With Incomplete Data

Performance Of Maximum Entropy Method And Self-Calibration

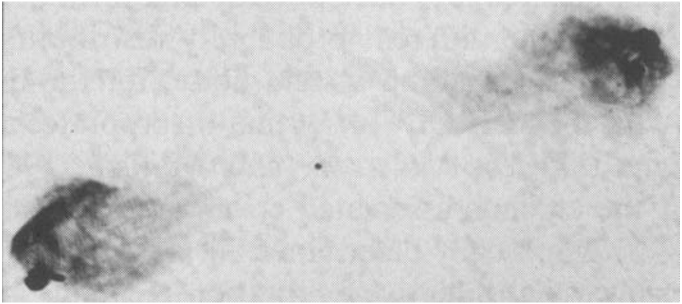


Figure: Three stages in the reduction of the observation of Cygnus A.

Mapping With Incomplete Data

Performance Of Maximum Entropy Method And Self-Calibration

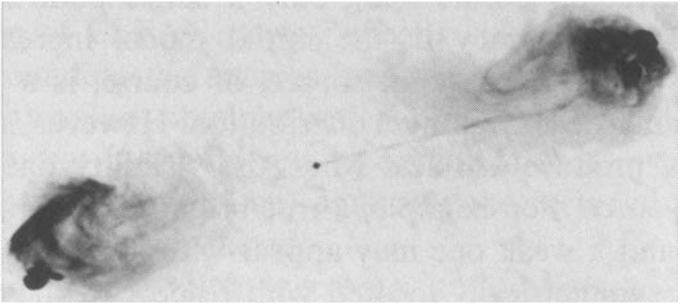


Figure: Three stages in the reduction of the observation of Cygnus A.

Outlook

- An important aspect of all electromagnetic radiation coming from outer space is its polarization.
- Studying the polarization of the radio signals gives valuable information about the kind of natural processes that lead to the radiation.
- Radio telescopes today have polarization receivers which record the four Stokes' parameter of the radiation. Different maps might then be made of different polarizations like linear circular etc.
- The decoding and analysis, as also the system response and digitization of this information forms an entire chapter of system design.

Acknowledgments

- I would like to thank Prof. R. K. Shevgaonkar for guiding me through this literature survey for my B. Tech. Seminar, and his constant encouragement and invaluable suggestions.
- I would also like to thank Prof. S. H. Patil for being my co-guide for this project.
- I would like to thank Prof. Vahia of Tata Institute of Fundamental Research for helping out with obtaining the literature for the survey.

Bibliography

- *Interferometry And Synthesis in Radio Astronomy*, A. Richard Thompson, James M. Moran, George W. Swenson, Jr., John Wiley & Sons, 1986.
- Cooper, B. F. C,. Correlators With Two-Bit Quantization, *Aust. J. Phys.*, **21**, 521-527, 1970.
- Cornwell, T. J., A Method Of Stabilizing The CLEAN Algorithm, *Astron. Astrophys.*, **121**, 281-285, 1983.
- Frater, R. H., Accurate Wideband Multiplier Square-Law Detector, *Rev. Sci. Instrum.*, **35**, 810-813, 1964.

Bibliography (cont.)

- Narayan, R. and R. Nityananda, Maximum Entropy Image Reconstruction - A Practical Non-Information-Theoretic Approach, *J. Astrophys. Astron.*, **3**, 419-450, 1982.
- Perley, R. A., J. W. Dreher And J.J. Cowan, The Jet And Filaments In Cygnus A, *Astrophys. J.*, **285**, L35-L38, 1984.