

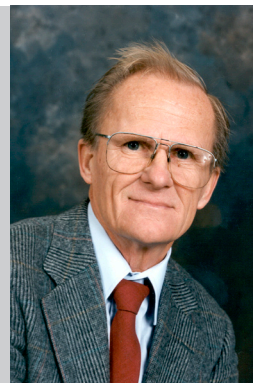


TAMPEREEN TEKNILLINEN YLIOPISTO  
TAMPERE UNIVERSITY OF TECHNOLOGY

*Tampere International Center for Signal Processing  
TICSP Series #38*

Peter Grünwald, Petri Myllymäki, Ioan Tabus, Marcelo Weinberger & Bin Yu (eds.)

## **Festschrift in Honor of Jorma Rissanen on the Occasion of his 75<sup>th</sup> Birthday**



Tampere International Center for Signal Processing.  
TICSP series # 38

Peter Grünwald, Petri Myllymäki, Ioan Tabus, Marcelo Weinberger &  
Bin Yu (eds.)

**Festschrift in Honor of Jorma Rissanen on the  
Occasion of his 75th Birthday**

Tampere International Center for Signal Processing  
Tampere 2008

© 2008 This publication follows TICSP copyright and licensing policy  
<http://ticsp.cs.tut.fi/index.php/TICSP> Copyright and License Agreement

ISBN 978-952-15-1962-8  
ISSN 1456-2774

Tampereen Yliopistopaino Oy, 2008

# Contents

<b>Preface</b>	vii
----------------	-----

## I Introduction

<b>1 A Conversation with Jorma Rissanen</b>	
<i>Pentti Huuhtanen, Erkki P. Liski, and Simo Puntanen</i>	
Department of Statistics, University of Tampere	1

## II Research Articles

<b>2 The MDL Principle, Penalized Likelihoods, and Statistical Risk</b>	
<i>Andrew R. Barron, Cong Huang, Jonathan Q. Li, and Xi Luo</i>	
Department of Statistics, Yale University	33
<b>3 Greedy and Relaxed Approximations to Model Selection: A Simulation Study</b>	
<i>Guilherme V. Rocha and Bin Yu</i>	
Statistics Department, University of California at Berkeley	63
<b>4 Experimental Design and Model Selection: The Example of Exoplanet Detection</b>	
<i>Vijay Balasubramanian, Klaus Larjo, and Ravi Sheth</i>	
David Rittenhouse Laboratories, University of Pennsylvania, and School of Natural Sciences, Institute for Advanced Study, Princeton	81
<b>5 Enumerative Coding for Tree Sources</b>	
<i>Álvaro Martín, Gadiel Seroussi, and Marcelo Weinberger</i>	
Universidad de la República, Uruguay and Hewlett-Packard Laboratories, Palo Alto, CA	93
<b>6 Stochastic Chains with Memory of Variable Length</b>	
<i>Antonio Galves and Eva Löcherbach</i>	
Universidade de São Paulo and Université Paris-Est	117
<b>7 Some Information-Theoretic Computations Related to the Distribution of Prime Numbers</b>	
<i>Ioannis Kontoyiannis</i>	
Department of Informatics, Athens University of Economics and Business, Athens, Greece	135

<b>8 MDL Model Averaging for Linear Regression</b>	
<i>Erkki P. Liski and Antti Liski</i>	
Department of Statistics, University of Tampere, and Department of Signal Processing, Tampere University of Technology	145
<b>9 What is Information?</b>	
<i>Jerzy Konorski and Wojciech Szpankowski</i>	
Gdansk University of Technology and Purdue University	155
<b>10 Genome Compression Using Normalized Maximum Likelihood Models for Constrained Markov Sources</b>	
<i>Ioan Tabus and Gergely Korodi</i>	
Department of Signal Processing, Tampere University of Technology	175
<b>11 Factorized NML Models</b>	
<i>Petri Myllymäki, Teemu Roos, Tomi Silander, Petri Kontkanen, and Henry Tirri</i>	
Complex Systems Computation Group, Helsinki Institute for Information Technology, Finland, and Nokia Research Center	189
<b>12 Towards the Multicomponent MDL Denoising</b>	
<i>Janne Ojanen, Jukka Heikkonen, and Kimmo Kaski</i>	
Laboratory of Computational Engineering, Helsinki University of Technology	205
<b>13 Context Adaptive Coding of Bi-level Images</b>	
<i>Søren Forchhammer</i>	
Institute for Communications, Technical University of Denmark	213
<b>14 Estimation of Sinusoidal Regression Models by Stochastic Complexity</b>	
<i>Ciprian Doru Giurcaneanu</i>	
Department of Signal Processing, Tampere University of Technology	229
<b>15 A Stochastic Complexity Perspective of Induction in Economics and Inference in Dynamics</b>	
<i>K. Vela Velupillai</i>	
Department of Economics, University of Trento, Italy and Girton College Cambridge	251
<b>16 Compression-based Methods for Nonparametric On-line Prediction, Regres- sion, Classification and Density Estimation of Time Series</b>	
<i>Boris Ryabko</i>	
Department of Applied Mathematics and Cybernetics, Siberian State University of Telecommunications and Computer Science	275

### III Personal Notes

<b>17 That Simple Device Already Used by Gauss</b>	
<i>Peter Grünwald</i>	
CWI Amsterdam, the Netherlands	293

<b>18 A Great Mind</b>	
<i>Paul Vitányi</i>	
CWI and University of Amsterdam, the Netherlands	305
<b>19 My Encounters with MDL</b>	
<i>Terry Speed</i>	
Department of Statistics at the University of California, Berkeley	311
<b>20 Jorma's Unintentional Contributions to the Source Coding Research in Eindhoven</b>	
<i>Tjalling Tjalkens and Frans Willems</i>	
Eindhoven University of Technology, the Netherlands	315



# Preface

This *Festschrift* is presented in honor of Jorma Rissanen, on the occasion of his 75<sup>th</sup> birthday on October 20th, 2007. It contains 20 contributions by Jorma's colleagues and friends covering a wide range of topics in the areas of information theory, statistical modeling and inference, data compression and applications of modeling in science, engineering, and economics, reflecting the many areas where Jorma's work had a great impact.

A special session will be organized in the IEEE Information Theory Workshop to be held in Porto, Portugal during May 5-9, 2008. A number of the contributions will be presented in the special session of the workshop and the final printed copy of the edited version of the present volume will be available.

Jorma's exemplary career and his discoveries are a continuous source of inspiration for generations of researchers and students. His discoveries in science have had a substantial impact on the foundations of statistics and information theory. They prove fruitful in a wide range of applications, including science (psychology, molecular biology, astrophysics), engineering (control, computer science, signal processing), and economics.

Jorma is energetic as ever, constantly producing novel, galvanizing contributions in his positions as professor emeritus at Tampere University of Technology, visiting professor at Helsinki University of Technology and University of London, and HIIT Fellow at Helsinki Institute for Information Technology.

This volume is built out of contributions continuing in many ways the line of thoughts and principles promoted by Jorma, giving a glimpse of the diversity of areas relevant to his research. This volume will also reveal some insights of his remarkable personal life. All those who have had the privilege to meet and work with him have been amazed to see in him such an example of staying straight in life and being faithful to his pursuit of truth and values through science.

We wish to Jorma the fulfillment of all his wishes and strength to continue transforming his wishes into reality.

Happy Birthday!

Bin, Ioan, Marcelo, Peter, and Petri

Berkely, Tampere, Palo Alto, Amsterdam, Helsinki

October 20th, 2007



## A Conversation with Jorma Rissanen

Pentti Huuhtanen, Erkki P. Liski and Simo Puntanen

### Brief Biography

Jorma Johannes Rissanen was born in Pielisjärvi, Finland, on October 20, 1932. He received his Master's degree in electrical engineering in 1956, his Licentiate and Doctor of Technology degrees in control theory and mathematics in 1960 and 1965, respectively, all from the Helsinki University of Technology, Finland. He joined IBM's San Jose Research Laboratory, California, 1966, where he remained for more than three decades, except for the academic year 1973–74, when he held the chair of control theory in Linköping University, Sweden. He is a fellow of the Helsinki Institute for Information Technology, University of Helsinki. He is also a Professor Emeritus at the Tampere University of Technology. Since 1998 he has been appointed Visiting Professor in Computer Science at Royal Holloway, University of London. He is a Foreign Member of the Finnish Academy of Science and Letters and he has received an Honorary Doctorate degree from the Tampere University of Technology, Finland, in 1992.

Jorma Rissanen is the founder of the Minimum Description Length (MDL) principle, a new inductive principle in statistical modeling. He has carried out significant research in the fields of control, prediction and system theories, relation theory, numerical mathematics, information and coding theory, probability theory and statistics. He has published more than a hundred research papers and the books *Stochastic Complexity in Statistical Inquiry* (1989, World Scientific) and *Information and Complexity in Statistical Modeling* (2007, Springer), and he holds 15 US patents. He continues to be active in research.

Jorma Rissanen has earned many honors and awards, including the 2006 Kolmogorov Medal of the Computer Learning Research Centre at Royal Holloway, University of London; an IEEE Information Society Golden Jubilee Award for Technological Innovation for the invention of arithmetic coding in 1998; the IEEE 1993 Richard W. Hamming medal 'For fundamental contributions to information theory, statistical inference, control theory, and the theory of complexity'; an IBM Corporate Award in 1991 for the MDL/PMDL principles and stochastic complexity; an IBM Outstanding Innovation Award in 1988 for work in statistical inference, information theory, and the theory of complexity; the Best Paper Award from the IEEE Information Theory Group in 1986 (covered all papers published in information theory during the preceding two-year period); the Best Paper Award from the International Federation of Automatic Control in 1981; an IBM Outstanding Innovation Award in 1980 for the introduction of arithmetic codes. He is an IEEE Fellow.

In 1956 he married Riitta Åberg, and they have a son Juhani and a daughter Natasha, one grandson Juhani and one granddaughter Elissa.

The following conversation took place both by email in the end of July and meeting in Tampere in August 2007.

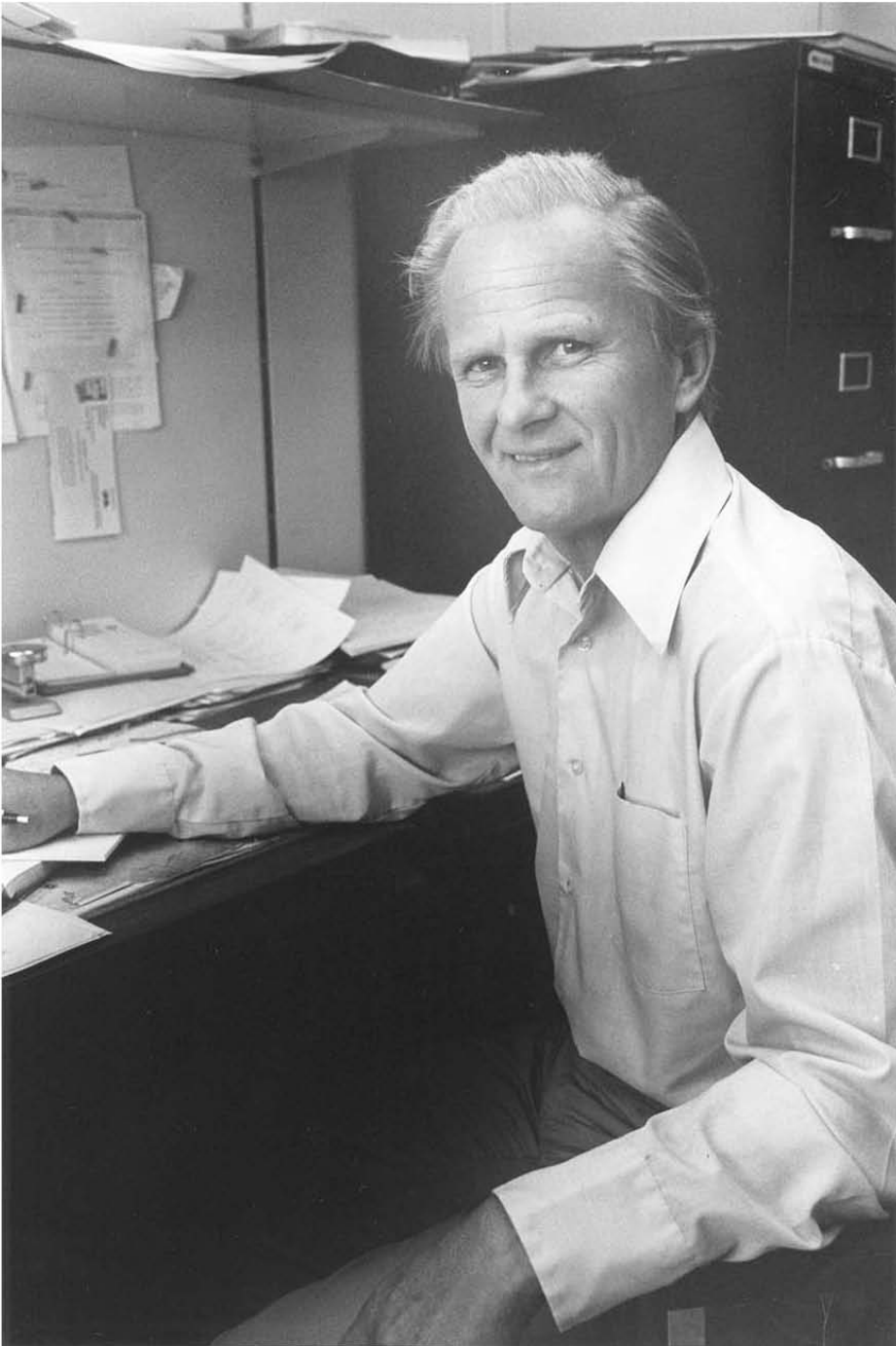


Figure 1: Jorma Rissanen in his San Jose office in 1985.

## Early life, Kemi, wartime, 1932–55

QUESTION: *Jorma, let us start at the very beginning. You were born on October 20, 1932 in Pielisjärvi, near then the Soviet border, but soon your family moved to Kemi, which is situated in the north-western part of Finland near to the Sweden–Finland border. Jorma, tell us something about your family and your childhood in Kemi.*

JORMA RISSANEN: There isn't all that much to tell. We lived in rented apartments in various suburbs of the small town, and just about all I remember of the period before the war was that it was cold and there were huge piles of snow.

*Did you have sisters or brothers?*

I had four sisters and a brother.

*Soon after you went to primary school in 1939 the Winter War broke out when the Soviet Union attacked Finland on November 30, 1939. The war lasted 100 days. What kind of time was that in your life?*

It did not affect us too much. Since Kemi is hundreds of kilometers from the Soviet border we had no bombing, even though we had to go to bomb shelters a few times.

*The period of peace after the Winter War was short. The Continuation War lasted from June 25, 1941 until September 19, 1944. Germany took part by providing critical material support and military cooperation to Finland and also some German troops operated in North Finland. When Finland in September 1944 made peace with the Soviet Union the so-called Lapland War broke out between Finland and Germany. Your home town Kemi was also in the middle of battles. Could you tell about your personal war experiences?*

Just before the fighting reached the house where we lived we took off to the woods together with the family of the owner of the house, a horse and a cow. There we waited for a few days until fighting passed the area. When the sound of the cannon shots seemed to come from farther away I decided to take a look at our home on the main road to north. A miles long German column of vehicles of any sort had been stopped on the road with dead soldiers with their arms lying nearby. I, together with the neighborhood boys, had a keen interest in guns and ammunition, and I thought that this is an excellent opportunity to get Schmeisser machine pistols. I picked two and hid them under the house where we lived. Every room of the apartment had bullet holes so that none of us would have survived if we had stayed. On the yard behind an Y-shape tree a dead German soldier with his machine rifle was lying. I was tempted to take the rifle but it was too long and heavy and I left it. Unfortunately I had no chance to try the machine pistols, because the next few days a Finnish soldier came to check if there are any weapons around, and my father told him that I had a cache under the house. I remember those as exciting times!

*How about your early schooling in wartime?*

After the Germans were pushed to Norway there was almost nothing to eat in Kemi, and I was sent with many others to Sweden, where I stayed one year without attending any school and missed the third grade. I learned Swedish though, which I however kept forgetting faster than what I was taught an hour every day from the third grade on in the junior high school.

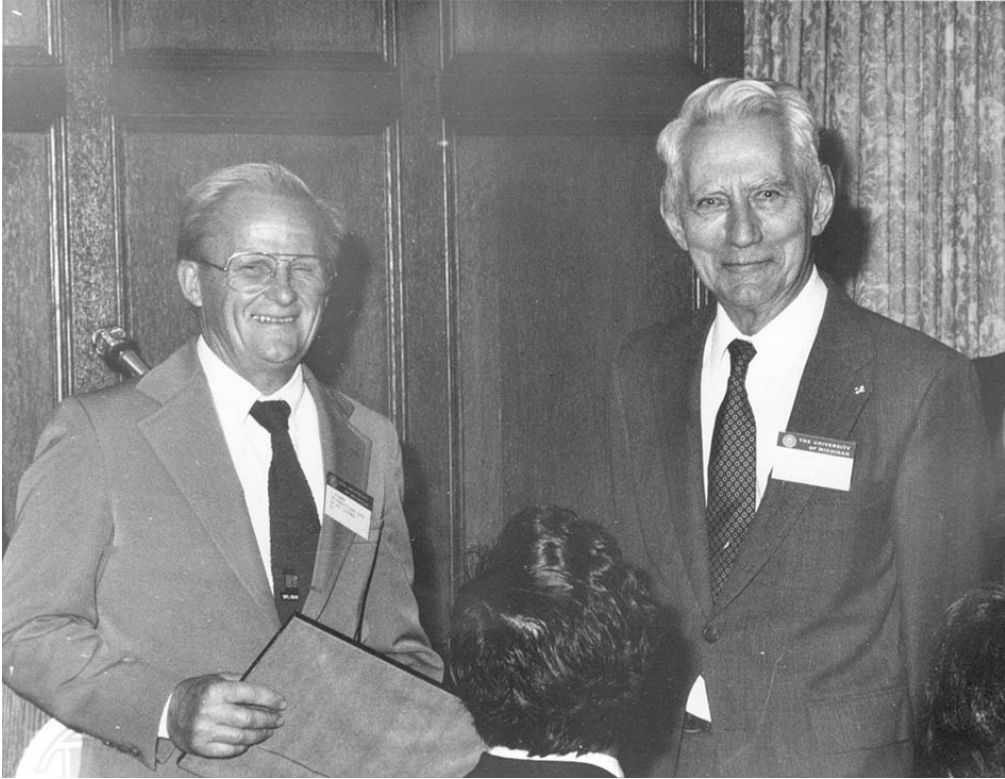


Figure 2: Conference in Ann Arbor, Michigan: Jorma Rissanen and Claude Shannon. Jorma Rissanen has received the Best Paper Award from IEEE Information Theory Group in 1986.

*What can you recall about your life during the elementary, junior high and high school days? Did you start showing any added aptitude in one subject over others at some stage?*

For a couple of years I had trouble with arithmetic and mathematics, but then all of a sudden, it seemed, I understood that arithmetic and mathematics is nothing but a game with certain rules, which you have to learn. After that it became my favorite field.

*Did you happen to meet inspiring teachers in science and mathematics?*

Our mathematics teacher happened to be excellent. He kept on reminding the students that there is nothing to learning mathematics and all of you are capable of getting at least the grade B. He left me with the long lasting impression that I should be able to learn anything if it is explained clearly.

Of science we were taught basic physics, the teacher of which was very good, too.

*How about the matriculation examination in the early 1950s? What kind of thoughts does that bring to your mind?*

The examination itself took one whole day. I remember the 3-hour mathematics test, which consisted of ten problems. One was an involved percentage type of problem, which was very hard

to decipher, and I missed it. However, since it was the only one I missed I still got the grade 'laudatur' and was allowed to attend the 2-week summer school for the Technical University of Helsinki. The other tests meant nothing to me and I must have passed them because I graduated.

## **Helsinki University of Technology, 1952–56**

*You enrolled in the undergraduate programme at Helsinki University of Technology (HUT) in 1952. How did you become interested in engineering studies?*

At the time there were two favorite ways to proceed towards higher education with good career prospects, the medical and technical fields, and my choice was clear: the most prestigious technical university of Finland was in Helsinki.

*After the years in Kemi, how did you like to live in Helsinki in your student years?*

I actually lived in Helsinki itself only the first year, after which I moved to Otaniemi and took the bus to the old technical university building, where all the lectures were given. I had no social life other than soccer, which we played just about every night in the newly built indoors arena in Otaniemi.

*Do you recall teachers or other people at HUT that you regard as influences on your career as it developed?*

Hans Blomberg, the professor of theoretical electrical engineering was excellent. So was the mathematics professor Kalle Väisälä, from whom I took all the courses he gave.

*You served in the army right after finishing the undergraduate degree in 1956. How did you cope with the mandatory military assignment?*

My military career ended prematurely when I hurt my knee and was operated, which gave me a good chance to study for the Licentiate degree in the military hospital.

*I recall that this knee problem has not been entirely solved and it has not allowed you to play soccer according to your talents? Is that right?*

That is right. I didn't play soccer in the following ten years until a knee specialist in California told me that there is nothing much wrong with the knee and go ahead and start playing, which I did.

*You started your Licentiate studies in 1957. What form were these studies? Was your aim already directed at a doctoral degree?*

There were no courses on control theory, and I picked the topic for graduate studies because of Hans Blomberg's interest in it. My aim certainly was a doctoral degree.

*You have told that one of your first places of employment was the Helsinki City Utility Company.*

It was my first steady employment; after all, I had just married and had to make a living.



Figure 3: Conference Dinner in Tampere Conference 1987: T. W. Anderson, Jorma Rissanen and Dorothy Anderson.

*So you were working and continuing your studies at the same time?*

Yes.

*When did your interests in mathematics and science begin to emerge?*

They grew gradually during the undergraduate studies.

*What was your dissertation about?*

It was about the horrible problem of adaptive control.

*Could you please be a bit more specific? Your answers makes me curious. . .*

Although I feel the less said about the thesis the better, let me just add that the most important and difficult part in problem selection is to pick a topic which both is tractable and reasonably significant. Since nobody had been able to do anything worthwhile about the adaptive control problem I should never have taken that as a PhD topic. Needless to say I wasted my time as well as that of the examiner, Olli Lokki, and produced absolutely nothing worthwhile.

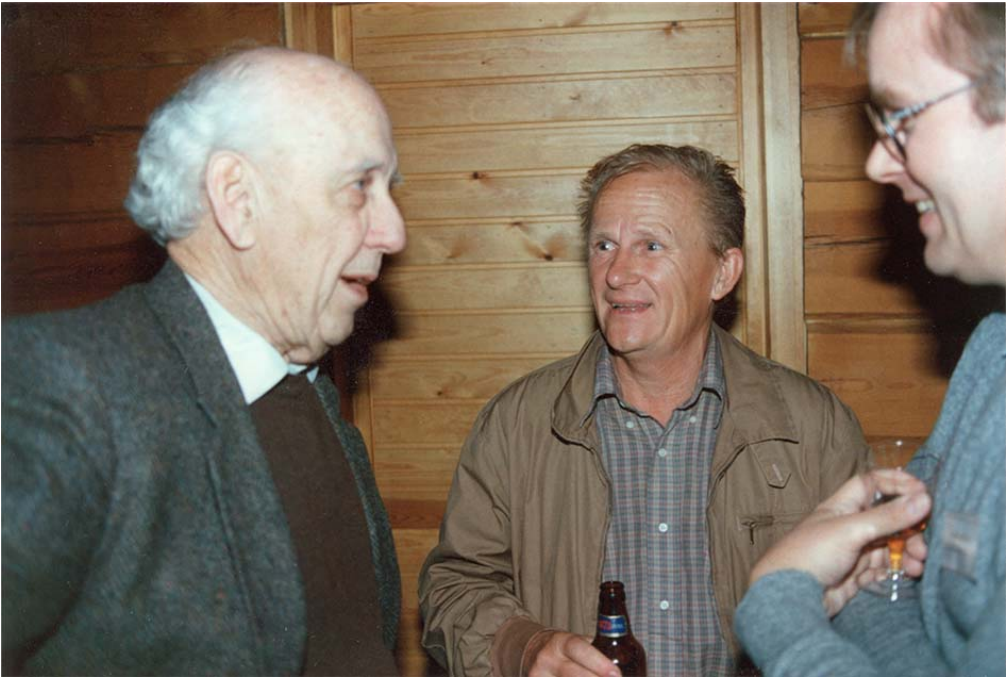


Figure 4: Sauna Party in Tampere Conference 1987: Ted Hannan, Jorma Rissanen, and Heikki Hella.

*Who was your thesis advisor and what kind of process it was to work for a doctoral degree at that time at HUT?*

My thesis advisor was Hans Blomberg, and since there were no formal lectures given on graduate level I simply started by reading literature given to me by him. I continued studying mathematics under the tutoring of Kalle Väisälä. The examination was done in my working out problems in books as home work – not avoiding the more difficult ones as instructed by Professor Väisälä.

*Often, work done for doctoral thesis shapes one's future conception of the field. Is this the case for you?*

No.

*Are there other people at HUT that you regard as influences on your career as it developed?*

No.

*You received the Doctor of Technology Degree in control theory and mathematics from the Helsinki University of Technology in 1965. Were there many other doctoral students in your field of research at HUT?*

Since I did most of my work away from the university while working for IBM, I didn't know of any other doctoral student in any field at HUT.

*Did you have any contacts with the Department of Mathematics at the University of Helsinki? At that time there were many world famous mathematicians like e.g. Rolf Nevanlinna, Paul Kustaanheimo, P. J. Myrberg, Gustav Elfving and Olli Lehto.*

I had no contact with the Department of Mathematics at the University of Helsinki, but I did know the impressive Paul Kustaanheimo and of course I knew Olli Lokki who was at the HUT. In fact, my first course in statistics were lectures given at HUT by Olli Lokki, who was the pioneer of statistics teaching in HUT.

*What were your plans for future after finishing your degree? How did you originally get interested in research career?*

It seems now that I have always regarded research as the only career for me, which was only strengthened while working in IBM Research.

*What about the teaching? Did you feel any call for teaching after finishing your degree?*

I didn't have a chance until later when I found out that teaching is too difficult to be left to the professionals.

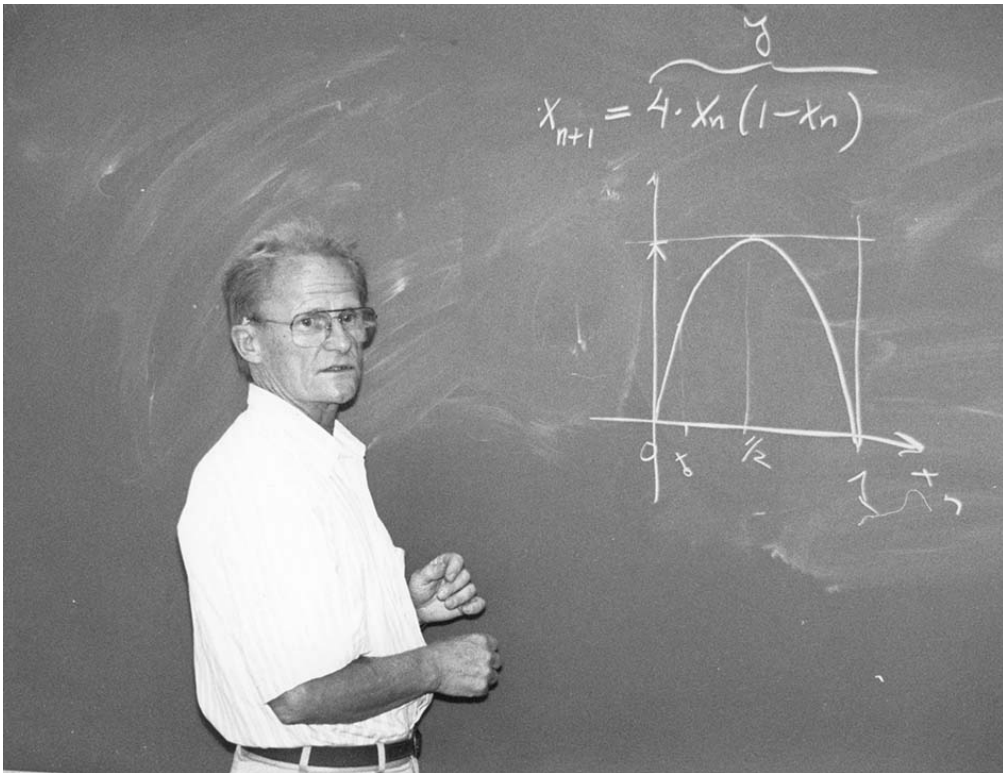


Figure 5: Workshop in Tampere 1990. Jorma Rissanen talking on "Stochastic Complexity in Linear Models".

## Sweden, 1958–64

*You started your international career in Sweden in 1958. You worked at ASEA in Västerås? Could you tell something about that time?*

I was told at the Helsinki City Utility Company that “by law we cannot fire you because of the military service, but don’t come back”. I never did. In fact I left the whole country and joined ASEA. I formulated and proved my very first theorem, one of the fundamental theorems in control (Rissanen 1961). It states that by a linear feedback you can move the poles wherever you wish but you cannot change their number.

*You moved to the IBM Nordic Laboratory in Stockholm in 1960.*

Yes, I did, and was sent for one year visit to Yorktown Heights and San Jose. These visits turned out to be crucial for my entire future.

*Information technology and industry lived their early stages in these times. Do you find that this has some effect on your orientation in research?*

The mission of the IBM Nordic Laboratories was industrial process control, but after a few years the company grew tired of waiting for economic results from that. The main problem was not control, which was easy if you knew what you were supposed to control. This turned my interests to modeling, and as it happens I never got out of it. The effect of information technology to my career was still some years away – and in fact came from an unexpected direction. But more of that a bit later.

*Your first published works were on linear systems and prediction theory. Are we right?*

Yes, I mentioned above my very first pole shifting theorem for linear feedback systems. In addition in the mid sixties I derived fast algorithms for factoring Hankel and Toeplitz matrices, which solve the so-called system identification problem as well as the Kalman type of prediction problem for ARMA processes (Rissanen 1973).

## To the USA and to IBM, 1964–

*You left Stockholm in 1964 and moved to the States. First to Electronic Associates in Princeton and Lockheed in Huntsville, but soon back to IBM’s San Jose Research Laboratory, California. That was a time of many changes of residence. Tell us what made you to get off the ground.*

Well, I needed a sponsor in the USA for emigration, and my boss in IBM USA during the visit had been promoted, which left me without a contact. I thought that who needs IBM and joined Electronic Associates. I realized however that maybe I had made a mistake and when IBM showed interest in getting us to San Jose we were just too happy to comply.

*Did you join a research group in San Jose or did you work more or less independently? Tell us about the research culture and atmosphere at IBM.*

I joined a research group but was given quite free hands to study and work on my own problems. This I took advantage of and became a professional student.

*San Jose is geographically pretty close to some of America's greatest universities. Did you develop connections with people in Stanford and Berkeley?*

Yes, indeed. I gave graduate lectures in Berkeley on prediction theory, and visited regularly professors Rudy Kalman and Tom Kailath in Stanford University. Later I actually was an adjunct professor in Stanford University.

*What kind of working habits did you follow? How would you describe a normal day?*

I started at about 8 AM and studied or worked on a problem leading to a paper until 5 PM. It took many years to get accustomed to such a regime without any obligations or directions – just study without focus and search for a meaningful problem. Every other day at noon time I played pickup soccer, which was a great distraction and broke the monotonous day.

*In 1973–74 you held the chair of control theory in Linköping University, Sweden. It was not your cup of tea?*

Immigrants often tend to feel nostalgia after a few years, and this happened to us as well. Also the lack of focus in my research had an effect, and when I saw the advertizement of a professor chair in Sweden I thought why not. It was a disastrous move. I found out that I don't like the field of control, I don't like to be a professor, and I don't like the climate nor the at that time socialistic Sweden. However, something happened, which maybe could not have happened otherwise. I was exposed to the exciting ideas of Chaitin, Kolmogorov, and Martin-Löf, which set my mind in fire. I found that this is what interests me, and finally I could focus on something that also could be of some interest to IBM. We returned to San Jose after just one year.

*Did you have any personal contact with Kolmogorov?*

I never had the honor of meeting him.

*Did you meet Harald Cramér while you were in Sweden?*

No, I never met him. But if you were to ask my statistical idols, I would definitely mention Harald Cramér and R. A. Fisher.

## **Arithmetic coding, stochastic complexity and MDL, 1975–**

*Many works you did at that time were somehow related to time series analysis. The Box–Jenkins approach to modeling ARIMA processes was described in a highly influential book by statisticians George Box and Gwilym Jenkins in 1970 (Box and Jenkins 1970). It seems that you already were familiar with these kinds of techniques at that time. Was it because of your system and prediction theory background?*

Yes indeed. The modeling problems for AR and ARMA processes were bread and butter in the control field a decade before the Box–Jenkins book.

*Later you had joint works with Ted Hannan who was one of the foremost experts in time series analysis. Would you like to say something about your collaborations with him?*

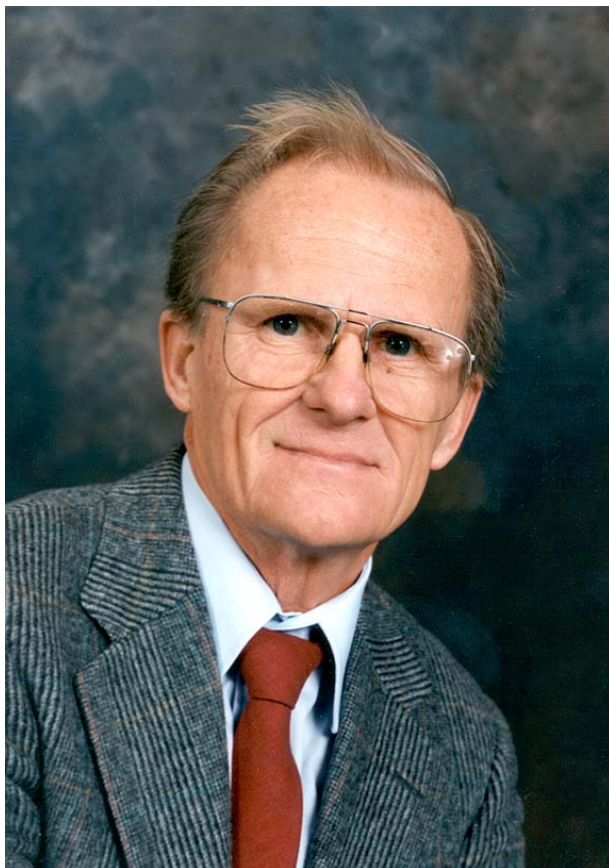


Figure 6: The IEEE announced that Jorma Rissanen is the recipient of the 1993 IEEE Richard W. Hamming Medal (IEEE Information Theory Society Newsletter 1993).

Ted loved to work on the difficult analysis problems arising in time series. His forte was analysis of existing problems rather than the creation of new ones. Among his many results was to derive a rather exact form for the penalty term in a criterion to find the number of parameters consistently. This also implies the consistency of the criterion BIC or the asymptotic approximation of the MDL criterion. To show something of my two-month visit to the Australian National University we had a joint *Biometrika* paper, written mostly by him, on estimation of ARMA order (Hannan and Rissanen 1982).

*In 1975 you introduced a new coding technique for data compression, called Arithmetic Coding, which is certainly one of your main contributions (Rissanen 1976). We may guess that coding theory was not one of your study subjects in the Helsinki University of Technology. Was it at IBM when you became interested in coding theory?*

I had never heard of the coding problem at the time I stayed at HUT. I got the idea of arithmetic coding from the brief paper by Kolmogorov, ‘Three Measures of Information’ (Kolmogorov 1965), during my stay as the professor of control theory in Sweden.

*How do you see the status of Arithmetic Coding nowadays?*

It is very simply the preferred way of doing coding for data compression. It has relegated the optimal, elegant, and then dominant Huffman codes to a graceful retirement – as someone put it.

*You introduced in 1983 a universal modeling algorithm Context, called now variable order Markov chains (Rissanen 1983). Tell us about the research that led to Context.*

I got the idea from the elegant data compression algorithm by Lempel and Ziv. Since I was advocating the view that all universal data compression algorithms must incorporate a model of the data, I first reinterpreted the main part in the LZ algorithm as a universal model in the huge class of ergodic processes. This then led to the universal modeling algorithm Context for the much smaller class of Markov chains. Because the class is smaller the model cost is smaller too, and you get a better compression if the data have any properties like the Markov chains.

*Many people would like to know what led you into formulating the principle that is now known as the MDL principle?*

Since with arithmetic coding you can encode any data modeled in any statistical manner in a completely mechanical and uniform way the key problem in data compression is to understand the statistical behavior of the data, which is modeling. I then turned the problem around and concluded that it is possible to measure the goodness of any model by its ability to compress the data. This relationship becomes evident by Kraft inequality, which establishes the logical equivalence between a distribution and the lengths of a (prefix) code tree.

*Was the MDL principle formulated the first time in your Automatica paper “Modeling by shortest data description” in 1978, or even before?*

Yes. The MDL principle as a concept was clear to me after the introduction of arithmetic coding in around 1975–1976, and in the *Automatica* paper I wanted to explain the principle to an audience not familiar with coding but one which I was familiar with, the control theory people.

*You were inspired by the Kolmogorov’s algorithmic theory of complexity. There are also the related works by Ray Solomonoff, G. J. Chaitin and Per Martin-Löf. Do you think that their works had some influence on your thinking?*

In a fundamental way. You see, it is one thing to understand things intuitively but quite different to see formally the relationships, which is what I learned from the writings of these distinguished men. In fact, there are still certain baffling both conceptual and technical issues, which I hope to be able to sort out some day.

*The idea of estimation-via-coding was presented in the computer science literature by Wallace and Boulton 1968. How is their approach related to MDL?*

First, their principle is expressed as minimizing the mean code length, which is then estimated. This results in a two-part code, which Wallace and his students have applied to a number of practical cases. The MDL principle on the other hand has been developed into a theory of inference rather than just a criterion for model selection.



Figure 7: Peter Grünwald and Jorma Rissanen during the 2002 IEEE Information Theory Workshop in Bangalore, India.

*Your paper ‘Stochastic complexity and modeling’ in Annals of Statistics 1986 extends certain fundamental results in information theory and statistics. Could you tell us in a few words about the significance of this contribution.*

At the time I wanted to prove that with a parametric family of distributions you cannot get a shorter code length for or, equivalently, a greater probability assigned to data sequences than a certain calculable bound – no matter how you construct the code. That is, such a bound, which I called stochastic complexity, is inherent for the family of distributions at hand. Later constructions were found with which the bound can be essentially achieved. This simply means that there are some absolute restrictions in modeling which you have to live with, and if your model gets away with such restrictions nobody can beat you. This is the aim with universal models.

*During the last three decades you have developed such ideas as stochastic complexity, universal model, and universal minimal sufficient statistics, which provide an information theoretic foundation for statistics. How do you assess the present status of your theory now?*

The universal minimal sufficient statistics in the algorithmic information theory, where a model of data is simply any finite set that includes the data, is due to Kolmogorov. When we deal with distributions as models, which must be estimated from data, there is a possibility to sharpen Kolmogorov’s result at least in two respects: First, the code length of a model, represented by the maximum likelihood estimate of its parameters, corresponds to the Kolmogorov complexity of the model. But the model can actually be represented more completely by the sequence of the maximum likelihood estimates, obtained from all the prefixes of the data sequence. This has no

counter part in the algorithmic theory, and clearly there is more information about the model in the sequence of the maximum likelihood estimates than in just the last estimate. Hence, the idea of ‘information’ that an estimated model represents is now different and more complex.

The second aspect which is missing in the algorithmic models is the idea of models that can be optimally ‘distinguished’ from a given set of data. I have written about how this provides a sense of optimality in hypothesis testing and greatly reduces the number of hypotheses we need to test.

My most recent work is of a new generation of universal models, which, while related to the predictive ways of constructing universal models like Dawid’s prequential ‘plug-in’ models or the equivalent predictive MDL models, are strictly better – provably so.

*If you were asked to list your three most important contributions, what would you list?*

The two most important contributions are undoubtedly Arithmetic Coding and the MDL principle. The third is tougher to pick. The theorem on the lower bound for the code length achievable for parametric families mentioned above together with the latest extensions of the MDL theory rank in my mind higher than Algorithm Context – perhaps because of the much greater difficulty in deriving them. In fact, I understand the proof in the *Annals of Statistics* paper only in my brighter moments.

## Philosophy

*In research, which is more important: conceptual foundations or technical perfection?*

I think the conceptual foundations are more important although their clarifications require often difficult analysis, which, moreover, tends to modify the initial concepts, and hence we cannot really separate the two.

*Do you see the computer as a tool in theoretical work?*

Computations of anything I have proposed have influenced my theoretical work (shown that I had overlooked something).

*What is your approach to a problem? Do you have certain techniques or would you say that is intuitive?*

My intuition amounts to seeing, or better feeling, what’s essential in the problem. Then by analysis comes a sharper isolation of that essential, which amounts to better understanding of the problem. The rest is detail – albeit sometimes crucial detail. An example is coding, in which the essential is just sorting and counting. However, the crucial thing is to understand exactly what it is that should be counted. In fact, this is also behind the entropy and the algorithmic information, so that in essence Kolmogorov’s three measures of information could be reduced to one. Clearly, to find the essence can be difficult, although in case of arithmetic coding it was not difficult, because of Kolmogorov’s paper, where the essence was spelled out. Nevertheless, after the invention of arithmetic codes my boss cum secretary exclaimed that he could not have invented that in a million years, or, more realistically, a million men could not find it in a year. In truth, the story of arithmetic coding is more involved than what I made it out here to be.



K. Vela Velupillai

Figure 8: Jorma Rissanen and Stefano Zambelli at the home of K. Vela Velupillai in Galway, Ireland, March 2005.

*What is your view on breadth versus depth in research? You have solved problems over a broad range and reached many deep findings. Some people tend to think that working deeply on a small topic is in contrast to having a wider interest.*

I see no contradiction in the two views. I sometimes tell young people that you need mathematics and deep analytic tools if you want to accomplish anything of lasting value; you cannot do that simply by being clever in the small.

In this context I might mention a nice poem by Piet Hein, which I saw in the office door of Terry Speed in Berkeley: Its name is Wide Road and it goes as follows:

To make a name for learning  
when other roads are barred,  
take something very easy  
and make it very hard.

*You have developed foundations of statistical modeling and carried out significant research in many fields like e.g. prediction and system theories, information and coding theory, computer learning, probability and statistics, and you have worked with computer scientists, engineers, information theorists, mathematicians, and statisticians. Could you give us your projections for statistical modeling in this new century?*

In my view statistics needs a solid foundation rather than just a collection of isolated techniques however clever.

It is not enough to claim to have found a method which works well or better than other methods on some data. It's even not enough to prove that the technique works on data generated by an imagined 'true' distribution. We need to understand why a technique works as it does, and why it is better than the competing approaches. This is what's missing in current statistics,

where all sorts of criteria for the model selection problem have been proposed. In addition, a sound statistical theory should be able to treat the estimation of both the parameter values and their number within a common theory. We ought to be able to formalize ideas like ‘information’, ‘complexity’, and ‘noise’, which I’m afraid can be done only with information theoretic means. As a specific example, there are good techniques for denoising. However, unless the idea of ‘noise’ is defined it becomes the part in the data that is removed. I hope that statistics will progress in a manner which makes sense and in which the fundamental concepts are defined and the limitations clearly understood.

*In many applications we are going to be faced with a lot of data generated by computational statistics or by a measuring device and then the theory has to keep up with it. Are we in a situation where there is more information than theory? Do you think there will be a revolution in statistical theory?*

I already explained above that the traditional statistics is unable to formalize the central concepts that are needed to capture properties in complex data, so that indeed there is more in the data than what the traditional statistical theory can explain. In my opinion, there will be drastic changes in statistics although one may suspect that the changes will be gradual, perhaps disguised as modifications of old approaches so that no foundation need be changed. Currently, foundations, such as they are, are ignored.

*Some writers seem to mix up MDL with Bayesian procedures. What is your view of the Bayesian philosophical framework?*

The MDL theory is based upon the MDL principle, which opens up an entirely different approach to statistics, free from the untenable assumption of a ‘true’ data generating distribution, while the Bayesian philosophy has no principle other than an unrestricted use of probabilities. Instead the central concept is the posterior distribution, whose interpretation is just as fuzzy as that of the prior. Moreover, since the prior affects the posterior its selection is crucial. In the MDL theory, where the use of priors is optional, their selection must be restricted so that they are encodable. This permits optimization of their selection, which cannot be done within the Bayesian philosophy, because nothing prevents you from putting it to unity on the data. The irresistible desire to peek into the data has created concepts like ‘empirical’ priors which clearly contradicts the very foundation of Bayesianism.

The usual confusion between the two approaches is understandable if one equates the MDL principle with the early criterion for model selection, also derived by Bayesian arguments and called BIC. Even though they are identical the MDL derivation attaches an asymptotic optimality to the criterion, while no such status can be given to BIC in terms of the Bayesian concepts. As a result the MDL criterion has been refined and developed further, while the BIC is a dead end.

Finally, the concepts like universal models, noise, statistical information, and complexity, which are central in the MDL theory, have no meaning in the Bayesian philosophy. It is true that Bayes’ formula creates a universal model, the so-called Bayesian mixture, which is good, but not because its goodness could be assessed by Bayesian means. Rather, it is good because it reaches the lower bound referred to above. Moreover, there are other universal models, which have not been found in the Bayesian philosophy and which have properties preferable to the Bayesian mixture. In summary, the MDL principle has created an entire theory with new concepts, which goes beyond any Bayesian technique and Bayesian philosophy whatever that is.

*What definition of probability do you use?*

Strictly speaking I use the probability that satisfies Kolmogorov's axioms. Sometimes, when it is preferable to talk about code length I mean by probability the number two raised to the negative power of the code length. After all, prefix codes are very concrete and there is nothing fuzzy about them. An example is the assignment of probability to a closed curve on the plane. It is easy to encode such curves by the chain link method, which then assigns a probability to the curve. Compare this with the horrendous task of defining a prior distribution for the set of all continuous closed curves on the plane.



Figure 9: Jorma Rissanen with Wojciech Szpankowski and Jacob Ziv at the ITA Workshop, San Diego, California, 2007.

## **Affiliations to Finland and current research**

*You were invited to “The 2nd International Tampere Conference in Statistics” in 1987. In consequence of this meeting we had the privilege to learn to know you personally. Since then you have been a regular visitor to Tampere. In 1987 Conference we had T. W. Anderson, George E. P. Box, C. R. Rao and Ted Hannan as keynote speakers. You surely had met all of them before?*

I do not recall having met Professor Box before that meeting. Actually, as you may know, I met Tarmo Pukkila, who was in charge for the Tampere Conference, first time in 1985 in Las Vegas in the ASA Conference – and there Tarmo invited me to Tampere in 1987.

*And now you even have a flat in Hervanta, Tampere. How do you share your time between California, Finland, and the numerous conference and lecturing trips?*

I spend about two months each year in Finland on three different time periods. The conference and lecture trips do not take that much of my time.

*You have been invited to numerous conferences, visited a number of universities, and have delivered many prestigious lectures, e.g. the Kolmogorov Lecture 2006. Could you mention some highlights?*

In addition to the Kolmogorov lecture I remember with pleasure a talk I gave in Norbert Wiener's 100 year memorial meeting in 1994, and the visit to Ann Arbor 1986, where Shannon himself handed me the best paper award. I also remember lectures in Beijing in 2005 as a guest of Microsoft, which provided a car with the driver for a week, and a memorial meeting of Z. C. Wei in Academia Sinica in Taipei 2005.

*From May 1995 until your mandatory retirement you held a part time professorship at the Tampere University of Technology. Tell us now how you came to TUT.*

TUT made me an honorary doctor in 1992, which followed with a part time professorship. When I had to retire from that TUT made other special arrangements, which allowed me to visit Tampere three times each year.

*Now every year you are teaching a course on statistical modeling at TUT, you are a fellow of the Helsinki Institute for Information Technology, and you have joint research projects in both places. Could you tell about these activities?*

I enjoy them greatly. Now I have access to peers and graduate students, which I never had in IBM Research.

*What do you want to say about your current research interests?*

I'm involved in applications of the MDL principle to practical problems. I'm also involved in theoretical work, some of which has been inspired by the applications.

*In addition to papers by only yourself you have plenty of papers with your collaborators. Can you shortly comment on the role of collaboration in your research?*

Although I have mostly worked alone collaboration has been quite important, in particular on arithmetic coding, whose practical implementation would have been outside of my skills.

*What about your PhD students? You surely must have a bunch of them?*

I have had access to PhD students mostly only in Finland, and I have found them both useful and invigorating. They tend to come up with unexpected questions, which expose embarrassing shortcomings in my original suggestions. Also, of course, their superb programming skills make the applications possible, which, in turn, create further problems.

## Personal

*You officially had to take the mandatory retirement from your part time professorship almost ten years ago. Do you consider yourself retired?*

I had no mandatory retirement from IBM Research. Rather, having already been involved in the institutions in Finland and London, I decided to retire from IBM five years ago and spend more time in Finland. I certainly do not consider myself retired. I often think that I should have retired from IBM earlier.

*What things do you like to do when you are not doing research? Tell us a little bit about your life outside of research and your hobbies or other activities.*

I must say that nowadays, when I don't play soccer any more, I have very little activities other than research. I walk my dog and work on the difficult to maintain yard. We live on a wild mountain side.

*You have held onto your passion for soccer. Tell us what kind of resonance soccer and bandy play have had for you.*

When I grew up in Kemi this game 'bandy', which is like soccer on ice with skates and a curved club to control a small ball, was my passion in winter and soccer in summer. I started to play recreational soccer in California in 1965 until my retirement in 2002. I would still play if there was an 'over 70 team' (no younger than 70 years old allowed in the team). Unfortunately, there does not seem to be enough players of that age.

*As far as we remember you enjoy reading mysteries. Who are your favorite writers?*

The British writers Ken Follet, Frederic Forsyth, Jack Higgins, and the Americans Clive Cussler, and Dan Brown.

*You did not learn English at school? What was your way of learning your perfect English?*

First, my command of English is far from perfect, but having lived in California for 40 years has helped.

*What does the future hold for you?*

I don't know.

*Yes indeed, we don't know what the future holds. We wish you well! Thanks very much Jorma for sharing your thoughts about your career and about science with us.*

I thank you for bothering to find out my thinking and making me feel as if I had done something worthwhile.



Figure 10: Jorma Rissanen looking at the sketch of a plot on sale in Kuru, Finland, April 2007.

## References

- G. E. P. Box, G. M. Jenkins (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, New York. [Third edition by G. E. P. Box, G. M. Jenkins and G. C. Reinsel, Prentice Hall, Englewood Cliffs, NJ, 1994.]
- E. J. Hannan, J. Rissanen (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69, 81–94.
- A. N. Kolmogorov (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1, 4–7.
- J. Rissanen (1961). Control system synthesis by analogue computer based on the “generalized linear feedback” concept. In *Analogue computation applied to the study of chemical processes, international seminar. Le calcul analogique appliqué à l’étude des processus chimiques, séminaire international. Actes/Proceedings*. (Robert Vichnevetsky, ed.), Gordon & Breach, New York.
- J. Rissanen (1973). Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrix polynomials. *Mathematics of Computation*, 27, 147–154.
- J. Rissanen (1976). Generalized Kraft inequality and arithmetic coding. *IBM Journal of Research and Development*, 20, 198–203.
- J. Rissanen (1978). Modeling by shortest data description. *Automatica*, 14, 465–471.
- J. Rissanen (1983). A universal data compression system. *IEEE Transactions on Information Theory*, 29, 656–664.
- J. Rissanen (1986). Stochastic complexity and modeling. *Annals of Statistics*, 14, 1080–1100.

- J. Rissanen (1989). *Stochastic Complexity in Statistical Inquiry*. World Scientific, Teaneck, NJ.
- J. Rissanen (2007). *Information and Complexity in Statistical Modeling*. Springer, New York.
- C. S. Wallace, D. M. Boulton (1968). An information measure for classification. *Computing Journal*, 11, 185–195.

## Acknowledgements

Special thanks go to Jarmo Niemelä for his assistance with L<sup>A</sup>T<sub>E</sub>X and for his help with the final editing of this interview, and in particular, for fixing up the list of Jorma Rissanen’s research publications. Thanks go also to Peter Grünwald, Virve Larmila, Teemu Roos, Terry Speed, Ioan Täbuş and K. Vela Velupillai for their help.

## Research publications of Jorma Rissanen

### Monographs and edited books

- [A1] Jorma Rissanen (1989). *Stochastic complexity in statistical inquiry*. World Scientific. (Series in Computer Science.) ISBN: 978-997150859-3.
- [A2] George Cybenko, Dianne P. O’Leary, Jorma Rissanen (eds.) (1999). *The mathematics of information coding, extraction, and distribution*. Proceedings of a workshop that was an integral part of the 1996–97 IMA program on Mathematics in high-performance computing. Univ. of Minnesota, Minneapolis, MN, USA, November 11–15, 1996. Springer. (The IMA Volumes in Mathematics and its Applications, vol. 107.) ISBN: 978-0-387-98665-4.
- [A3] Jorma Rissanen (2007). *Information and complexity in statistical modeling*. Springer. (Series: Information Science and Statistics.) ISBN: 978-0-387-36610-4.

### Articles in refereed journals and collections

- [B1] J. Rissanen (1961). Control system synthesis by analogue computer based on the “generalized linear feedback” concept. In *Analogue computation applied to the study of chemical processes, international seminar. Le calcul analogique appliqué à l’étude des processus chimiques, séminaire international. Actes/Proceedings*. (Robert Vichnevetsky, ed.), Gordon & Breach, New York.
- [B2] J. Rissanen (1963). On the theory of self-adjusting models. *Automatica*, 1 (4), 297–309.
- [B3] J. Rissanen (1966). Performance deterioration of optimum systems. *IEEE Transactions on Automatic Control*, 11 (3), 530–532.
- [B4] N. McClamroch, J. Rissanen (1967). A result on the performance deterioration of optimum systems. *IEEE Transactions on Automatic Control*, 12 (2), 209–210.
- [B5] J. Rissanen (1967). On duality without convexity. *Journal of Mathematical Analysis and Applications*, 18 (2), 269–275.
- [B6] A. Chang, J. Rissanen (1968). Regulation of incompletely identified linear systems. *SIAM Journal on Control and Optimization*, 6 (3), 327–348.
- [B7] J. Rissanen, L. Barbosa (1969). A factorization problem and the problem of predicting non-stationary vector-valued stochastic processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 12, 255–266.
- [B8] J. Rissanen, L. Barbosa (1969). Properties of infinite covariance matrices and stability of optimum predictors. *Information Sciences*, 1 (3), 221–236.

- [B9] J. Rissanen (1970). On factoring positive operators. *Journal of Mathematical Analysis and Applications*, 32 (3), 505–511.
- [B10] J. Rissanen (1970). Reproof of a theorem on duality without convexity. *Journal of Mathematical Analysis and Applications*, 29 (2), 429–431.
- [B11] Thomas Kailath, J. Rissanen (1971). The stochastic realization problem. In *Proceedings of the 1971 IEEE Conference on Decision and Control including the 10th Symposium on Adaptive Processes, Miami Beach, FL, 15–17 December 1971*, vol. 10, p. 546.
- [B12] J. Rissanen (1971). Maximum power feedback laws. *International Journal of Control*, I. Ser., 14, 233–240.
- [B13] J. Rissanen (1971). On optimum root-finding algorithms. *Journal of Mathematical Analysis and Applications*, 36 (1), 220–225.
- [B14] J. Rissanen (1971). Recursive identification of linear systems. *SIAM Journal on Control and Optimization*, 9, 420–430.
- [B15] J. Rissanen (1972). Recursive evaluation of Padé approximants for matrix sequences. *IBM Journal of Research and Development*, 16, 401–406.
- [B16] J. Rissanen, T. Kailath (1972). Partial realization of random systems. *Automatica*, 8 (4), 389–396.
- [B17] J. Rissanen, T. Kailath (1972). Partial realization of random systems. In *Proceedings of the IFAC Fifth World Congress (Paris, 1972), Part 4: Education, feedback, regulators, linear and non-linear systems; Identification, differential games, discrete and stochastic systems, Paper No. 35.6*, International Federation of Automatic Control, Düsseldorf, p. 6.
- [B18] J. Rissanen (1973). Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrix polynomials. *Mathematics of Computation*, 27, 147–154.
- [B19] J. Rissanen (1973). Bounds for weight balanced trees. *IBM Journal of Research and Development*, 17, 101–105.
- [B20] J. Rissanen (1973). A fast algorithm for optimum linear predictors. *IEEE Transactions on Automatic Control*, 18 (5), 555.
- [B21] P. E. Caines, J. Rissanen (1974). Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes (corresp.). *IEEE Transactions on Information Theory*, 20 (1), 102–104.
- [B22] J. Rissanen (1974). Basis of invariants and canonical forms for linear dynamic systems. *Automatica*, 10, 175–182.
- [B23] Jorma Rissanen (1974). Solution of linear equations with Hankel and Toeplitz matrices. *Numerische Mathematik*, 22, 361–366.
- [B24] J. Rissanen, Bostwick F. Wyman (1975). Duals of input/output maps. In *Category theory applied to computation and control (Proc. First Internat. Sympos., San Francisco, Calif., February 25–26, 1974), Lecture Notes in Computer Science*, vol. 25, Springer, Berlin, pp. 204–208.
- [B25] Jorma Rissanen (1975). Canonical Markovian representations and linear prediction. In *Proceedings of the IFAC 6th World Congress (Boston/Cambridge, Mass., 1975), Part 1, Paper No. 29.3*, International Federation of Automatic Control, Düsseldorf, p. 9.
- [B26] J. Rissanen (1976). Minimax entropy estimation of models for vector processes. In *System identification: Advances and Case Studies* (D. G. Lainiotis, R. K. Mehra, eds.), *Mathematics in Science and Engineering*, vol. 126, Academic Press, New York, pp. 97–117.
- [B27] J. Rissanen, L. Ljung (1976). Estimation of optimum structures and parameters for linear systems. In *Mathematical systems theory: Proceedings of the international symposium held in Udine, Italy, June 16–27, 1975, Lecture Notes in Economics and Mathematical Systems*, vol. 131, Springer, Berlin, pp. 92–110.

- [B28] J. J. Rissanen (1976). Generalized Kraft inequality and arithmetic coding. *IBM Journal of Research and Development*, 20 (3), 198–203.
- [B29] Jorma Rissanen (1976). System estimation by entropy criterion. In *Proceedings of the Ninth Hawaii International Conference on System Sciences (Univ. Hawaii, Honolulu, Hawaii, 1976)*, Western Periodicals, North Hollywood, Calif., pp. 199–200.
- [B30] Jorma Rissanen (1977). Independent components of relations. *ACM Transactions on Database Systems*, 2 (4), 317–325.
- [B31] Lennart Ljung, Jorma Rissanen (1978). On canonical forms, parameter identifiability and the concept of complexity. In *Proceedings of the 4th IFAC Symposium on Identification and system parameter estimation, Tbilisi, USSR, 1976, Part 3* (N. S. Rajbman, ed.), North-Holland, Amsterdam, pp. 1415–1426.
- [B32] J. Rissanen (1978). Modeling by shortest data description. *Automatica*, 14, 465–471.
- [B33] J. Rissanen (1978). Theory of relations for databases – a tutorial survey. In *Mathematical foundations of computer science 1978: Proceedings, 7th Symposium, Zakopane, Poland, 1978* (Józef Winkowski, ed.), *Lecture Notes in Computer Science*, vol. 64, Springer-Verlag, Berlin, pp. 537–551.
- [B34] Jorma Rissanen (1978). Minimax codes for finite alphabets (corresp.). *IEEE Transactions on Information Theory*, 24 (3), 389–392.
- [B35] J. Rissanen (1979). Shortest data description and consistency of order estimates in ARMA-processes. In *International Symposium on Systems Optimization and Analysis (Rocquencourt, 1978)*, *Lecture Notes in Control and Information Sciences*, vol. 14, Springer, Berlin, pp. 92–98.
- [B36] J. Rissanen, P. E. Caines (1979). The strong consistency of maximum likelihood estimators for ARMA processes. *The Annals of Statistics*, 7 (2), 297–315.
- [B37] J. Rissanen, G. G. Langdon, Jr. (1979). Arithmetic coding. *IBM Journal of Research and Development*, 23 (2), 149–162.
- [B38] Jorma Rissanen (1979). Arithmetic codings as number representations. *Acta Polytechnica Scandinavica, Mathematics and Computer Science Series*, 31, 44–51.
- [B39] J. Rissanen (1980). Consistent order estimates of autoregressive processes by shortest description of data. In *Analysis and optimisation of stochastic systems. Based on the Proceedings of the International Conference held at the University of Oxford from 6–8 September, 1978, organised by The Institute of Mathematics and its Applications* (O. Jacobs, M. Davis, M. Dempster, C. Harris, P. Parks, eds.), Academic Press, New York.
- [B40] G. Langdon, J. Rissanen (1981). Compression of black-white images with arithmetic coding. *IEEE Transactions on Communications*, 29 (6), 858–867.
- [B41] Jorma Rissanen (1981). Order estimation in Box-Jenkins model for time series. *Methods of Operations Research*, 44, 143–150.
- [B42] Jorma Rissanen, Glen G. Langdon, Jr. (1981). Universal modeling and coding. *IEEE Transactions on Information Theory*, 27 (1), 12–23.
- [B43] E. J. Hannan, J. Rissanen (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika*, 69 (1), 81–94.
- [B44] Glen G. Langdon, Jr., Jorma Rissanen (1982). A simple general binary source code. *IEEE Transactions on Information Theory*, 28 (5), 800–803.
- [B45] J. Rissanen (1982). Estimation of structure by minimum description length. *Circuits, Systems, and Signal Processing*, 1 (3–4), 395–406.
- [B46] Jorma Rissanen (1982). On equivalences of database schemes. In *Proceedings of the ACM symposium on principles of database systems, March 29–31, 1982, Los Angeles, California*, Association for Computing Machinery, pp. 23–26.

- [B47] Jorma Rissanen (1982). Tight lower bounds for optimum code length. *IEEE Transactions on Information Theory*, 28 (2), 348–349.
- [B48] E. J. Hannan, J. Rissanen (1983). Errata: “Recursive estimation of mixed autoregressive-moving average order” [Biometrika 69 (1), 81–94, 1982; MR0655673 (84e:62136)]. *Biometrika*, 70 (1), 303.
- [B49] Glen G. Langdon, Jr., Jorma Rissanen (1983). Correction to “A simple general binary source code” (Sep 1982, 800–803). *IEEE Transactions on Information Theory*, 29 (5), 778–779.
- [B50] Glen G. Langdon, Jr., Jorma J. Rissanen (1983). A double-adaptive file compression algorithm. *IEEE Transactions on Communications*, 31 (11), 1253–1255.
- [B51] J. Rissanen (1983). Probability estimation for symbols – observed or not. In *International Symposium on information theory held at Quebec, Canada on September 26–30, 1983. Abstracts of papers*, Institute of Electrical and Electronics Engineers Inc., New York.
- [B52] Jorma Rissanen (1983). Information in prediction and estimation. In *Proceedings of the 22nd IEEE Conference on Decision and Control, San Antonio, TX, 14–16 December 1983*, vol. 22, pp. 308–310.
- [B53] Jorma Rissanen (1983). A universal data compression system. *IEEE Transactions on Information Theory*, 29 (5), 656–664.
- [B54] Jorma Rissanen (1983). A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11 (2), 416–431.
- [B55] K. Mohiuddin, J. J. Rissanen, R. Arps (1984). Lossless binary image compression based on pattern matching. In *Proceedings of the International Conference on Computers, Systems, and Signal Processing, December 1984, Bangalore, India*, pp. 447–451.
- [B56] Jorma Rissanen (1984). Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory*, 30 (4), 629–636.
- [B57] J. Rissanen (1985). Minimum description length principle. In *Encyclopedia of Statistical Sciences* (S. Kotz, N. L. Johnson, eds.), vol. 5, John Wiley and Sons, pp. 523–527.
- [B58] J. Rissanen, V. Wertz (1985). Structure estimation by accumulated prediction error criterion. In *Proceedings of the 7th IFAC/IFORS Symposium on Identification and System Parameter Estimation held in 3–7 July 1985 in York, England* (H. A. Barker, Young P. C., eds.), Pergamon Press.
- [B59] Stephen Todd, Glen G. Langdon, Jr., Jorma Rissanen (1985). Parameter reduction and context selection for compression of gray-scale images. *IBM Journal of Research and Development*, 29 (2), 188–193.
- [B60] L. Gerencsér, J. Rissanen (1986). A prediction bound for Gaussian ARMA processes. In *Proceedings of the 25th IEEE Conference on Decision and Control, Athens, Greece*, vol. 3, pp. 1487–1490.
- [B61] Jorma Rissanen (1986). Complexity of strings in the class of Markov sources. *IEEE Transactions on Information Theory*, 32 (4), 526–532.
- [B62] Jorma Rissanen (1986). Order estimation by accumulated prediction errors. *Journal of Applied Probability*, Special vol. 23A: Essays in time series and allied processes, 55–61.
- [B63] Jorma Rissanen (1986). Predictive and nonpredictive minimum description length principles. In *Time series and linear systems* (S. Bittanti, ed.), *Lecture Notes in Control and Information Sciences*, vol. 86, Springer, Berlin, pp. viii, 115–140.
- [B64] Jorma Rissanen (1986). A predictive least squares principle. *IMA Journal of Mathematical Control and Information*, 3 (2–3), 211–222. Special issue: Parametrization problems.
- [B65] Jorma Rissanen (1986). Stochastic complexity and modeling. *The Annals of Statistics*, 14 (3), 1080–1100.

- [B66] Jorma Rissanen (1986). Stochastic complexity and statistical inference. In *Analysis and optimization of systems (Antibes, France, 1986)*, *Lecture Notes in Control and Information Sciences*, vol. 83, Springer, Berlin, pp. 393–407.
- [B67] J. Rissanen (1987). Complexity and information in contingency tables. In *Proceedings of The Second International Tampere Conference in Statistics, June 1–4, 1987, Tampere, Finland* (Tarmo Pukkila, Simo Puntanen, eds.), Report A 184, University of Tampere, Department of Mathematical Sciences, Tampere, Finland. Invited paper.
- [B68] Jorma Rissanen (1987). Comment on “Rational transfer function approximation” by E. J. Hannan. *Statistical Science*, 2 (2), 156–157.
- [B69] Jorma Rissanen (1987). Stochastic complexity and the MDL principle. *Econometric Reviews*, 6 (1), 85–102.
- [B70] Jorma Rissanen (1987). Stochastic complexity (with discussion). *Journal of the Royal Statistical Society. Series B. Methodological*, 49 (3), 223–239, 253–265.
- [B71] Jorma Rissanen, Mati Wax (1987). Measures of mutual and causal dependence between two time series. *IEEE Transactions on Information Theory*, 33 (4), 598–601.
- [B72] E. J. Hannan, J. Rissanen (1988). The width of a spectral window. *Journal of Applied Probability*, Special vol. 25A: A celebration of applied probability, 301–307.
- [B73] Jorma Rissanen (1988). Comment on “ARMA memory index modeling of economic time series” by H. J. Bierens. *Econometric Theory*, 4 (1), 61–64.
- [B74] Jorma Rissanen (1988). Stochastic complexity and the maximum entropy principle. In *Maximum-entropy and Bayesian methods in science and engineering, vol. 1 (Laramie, WY, 1985, and Seattle, WA, 1986/1987)*, *Fundamental Theories of Physics*, Kluwer Acad. Publ., Dordrecht, pp. 161–171.
- [B75] J. Rissanen, K. M. Mohiuddin (1989). A multiplication-free multialphabet arithmetic code. *IEEE Transactions on Communications*, 37 (2), 93–98.
- [B76] J. Rissanen (1990). Complexity of models. In *Complexity, Entropy and the Physics of Information: Proceedings of the Santa Fe Institute Workshop, May 29 to June 10, 1989* (Wojciech H. Zurek, ed.), *Santa Fe Institute studies in the sciences of complexity*, vol. 8, Westview Press, pp. 117–126.
- [B77] J. Rissanen (1992). Discussion of “Prequential analysis, stochastic complexity and Bayesian inference” by A. P. Dawid. In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting* (J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith, eds.), Oxford University Press, Oxford, pp. 121–122.
- [B78] Jorma Rissanen, Terry P. Speed, Bin Yu (1992). Density estimation by stochastic complexity. *IEEE Transactions on Information Theory*, 38 (2), 315–323.
- [B79] László Gerencsér, Jorma Rissanen (1993). Asymptotics of predictive stochastic complexity. In *New directions in time series analysis, Part II. Proceedings of the 1990 IMA Workshop* (D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, M. S. Taquq, eds.), *IMA Volumes in Mathematics and its Applications*, vol. 46, Springer, New York, pp. 93–112.
- [B80] J. Rissanen (1993). The minimal description length principle. In *Encyclopedia of Microcomputers* (Allen Kent, James G. Williams, eds.), vol. 11, Marcel Dekker, New York, NY, pp. 151–158.
- [B81] M. J. Weinberger, M. Feder, J. Rissanen (1993). Sequential model estimation for universal coding and the predictive stochastic complexity of finite-state sources. In *Proceedings of IEEE International Symposium on Information Theory, San Antonio, TX, January 17–22, 1993*, Institute of Electrical and Electronics Engineers, inc, Piscataway, NJ, p. 52.
- [B82] L. Gerencsér, J. H. van Schuppen, J. Rissanen, Z. Vágó (1994). Stochastic complexity, selftuning and optimality. In *Proceedings of the 33rd IEEE Conference on Decision and Control, Lake Buena Vista, FL, 14–16 December 1994*, vol. 1, pp. 652–654.

- [B83] J. Rissanen (1994). Noise separation and MDL modeling of chaotic processes. In *From Statistical Physics to Statistical Inference and Back. Proceedings of NATO Summer School* (P. Grassberger, J.-P. Nadal, eds.), NATO ASI Series, Kluwer Academic Publishers, pp. 317–330.
- [B84] Jorma Rissanen, Eric Sven Ristad (1994). Language acquisition in the MDL framework. In *Language computation. DIMACS workshop on human language, March 20–22, 1992, at Princeton Univ., Princeton, NJ, USA* (Eric Sven Ristad, ed.), *DIMACS Ser. Discrete Math. Theor. Comput. Sci.*, vol. 17, American Mathematical Society, Providence, RI, pp. 149–166.
- [B85] Jorma Rissanen, Eric Sven Ristad (1994). Unsupervised classification with stochastic complexity. In *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: an Informational Approach (Knoxville, TN, 1992)* (Hamparsum Bozdogan, ed.), vol. 2, Kluwer Academic Publ., Dordrecht, pp. 14–15, 171–182.
- [B86] Søren Forchhammer, Jorma Rissanen (1995). Coding with partially hidden Markov models. In *Proceedings of the IEEE Data Compression Conference, DCC 1995, Snowbird, Utah, March 28–30, 1995* (James A. Storer, Martin Cohn, eds.), IEEE Computer Society Press, pp. 92–101.
- [B87] L. Gerencsér, J. H. van Schuppen, J. Rissanen, Z. Vágó (1995). Parametric uncertainty and control performance in stochastic adaptive control. In *Proceedings of the 5th IFAC Symposium on Adaptive Systems in Control and Signal Processing (ACASP'95), Budapest, Hungary, 1995*, pp. 79–82.
- [B88] Manish Mehta, Jorma Rissanen, Rakesh Agrawal (1995). MDL-based decision tree pruning. In *Proceedings of the First International Conference on Knowledge Discovery in Databases and Data Mining (KDD-95), Montreal, Canada, August 20–21, 1995*, pp. 216–221.
- [B89] J. Rissanen (1995). Stochastic complexity and its applications. In *Workshop on Model Uncertainty and Model Robustness, June 30th to July 2nd 1995, Bath, England*. Electronic publication. Available from <http://www-old.stat.duke.edu/conferences/bath/rissanen.ps>.
- [B90] Jorma Rissanen (1995). Stochastic complexity in learning. In *Computational Learning Theory, Second European Conference, EuroCOLT '95, Barcelona, Spain, March 1995, Proceedings* (Paul M. B. Vitányi, ed.), *Lecture Notes in Artificial Intelligence*, vol. 904, Springer, pp. 196–210. Invited lecture.
- [B91] Marcelo J. Weinberger, Jorma Rissanen, Meir Feder (1995). A universal finite memory source. *IEEE Transactions on Information Theory*, 41 (3), 643–652.
- [B92] Søren Forchhammer, Jorma Rissanen (1996). Partially hidden Markov models. *IEEE Transactions on Information Theory*, 42 (4), 1253–1256.
- [B93] Manish Mehta, Rakesh Agrawal, Jorma Rissanen (1996). Sliq: a fast scalable classifier for data mining. In *Advances in Database Technology – EDBT'96, 5th International Conference on Extending Database Technology, Avignon, France, March 25–29, 1996, Proceedings* (Peter M. G. Apers, Mokrane Bouzeghoub, Georges Gardarin, eds.), *Lecture Notes in Computer Science*, vol. 1057, Springer, pp. 18–32.
- [B94] J. Rissanen (1996). Information theory and neural nets. In *Mathematical Perspectives in Neural Networks* (P. Smolensky, M. C. Mozer, D. E. Rumelhart, eds.), Lawrence Erlbaum Associates, Mahwah, NJ, chapter 16, pp. 567–602.
- [B95] J. Rissanen (1996). Stochastic complexity – an introduction. In *Computational Learning and Probabilistic Reasoning*, Wiley, Chichester, England, chapter 2, pp. 33–41.
- [B96] J. Rissanen, Bin Yu (1996). Learning by MDL. In *Learning and Geometry: Computational Approaches* (David W. Kueker, Carl H. Smith, eds.), *Progress in Computer Science and Applied Logic*, vol. 14, Birkhäuser, Boston, MA, pp. 3–19.
- [B97] Jorma J. Rissanen (1996). Fisher information and stochastic complexity. *IEEE Transactions on Information Theory*, 42 (1), 40–47.

- [B98] Marcelo J. Weinberger, Jorma Rissanen, Ronald Arps (1996). Applications of universal context modeling to lossless compression of gray-scale images. *IEEE Transactions on Image Processing*, 5 (4), 575–586.
- [B99] J. Rissanen (1997). Shannon-Wiener information and stochastic complexity. In *Proceedings of the Norbert Wiener Centenary Congress, 1994 (East Lansing, MI, November 27–December 3, 1994)* (V. Mandrekar, P. R. Masani, eds.), *Proceedings of Symposia in Applied Mathematics*, vol. 52, American Mathematical Society, Providence, RI, pp. 331–341.
- [B100] J. Rissanen (1997). Stochastic complexity in learning. *Journal of Computer and System Sciences*, 55 (1, part 2), 89–95. Second Annual European Conference on Computational Learning Theory (EuroCOLT '95), Barcelona, 1995.
- [B101] Ioan Tăbuș, Jorma Rissanen, Jaakko Astola (1997). Adaptive L-predictors based on finite state machine context selection. In *Proceedings of ICIP'97 International Conference on Image Processing, Santa Barbara, California, October 1997*, pp. 401–404.
- [B102] Andrew Barron, Jorma Rissanen, Bin Yu (1998). The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44 (6), 2743–2760. Special issue: Information Theory: 50 Years of Discovery.
- [B103] J. Rissanen, G. Shedler (1998). Failure-time prediction. *Journal of Statistical Planning and Inference*, 66 (2), 193–210.
- [B104] I. Tăbuș, C. Popeea, J. Rissanen, Astola J. (1998). Alphabet extensions for Markov sources. In *Proceedings of ITW'98 Information Theory Workshop, San Diego, California, February 8–11, 1998*, p. 78.
- [B105] J. Rissanen (1999). MDL denoising with wavelets. In *IEEE Information Theory Workshop on Detection, Estimation, Classification and Imaging (DECI), February 24–26, 1999, Santa Fe, NM*. Electronic publication. Available from <http://www.ifp.uiuc.edu/itw-deci/psfiles/rissanen.ps>.
- [B106] Jorma Rissanen (1999). Discussion of paper “Minimum message length and Kolmogorov complexity” by C. S. Wallace and D. L. Dowe. *The Computer Journal*, 42 (4), 327–329.
- [B107] Jorma Rissanen (1999). Fast universal coding with context models. *IEEE Transactions on Information Theory*, 45 (4), 1065–1071.
- [B108] Jorma Rissanen (1999). Hypothesis selection and testing by the MDL principle. *The Computer Journal*, 42 (4), 260–269. Invited paper for the special issue devoted to Kolmogorov complexity.
- [B109] Jorma Rissanen (1999). Rejoinder. *The Computer Journal*, 42 (4), 343–344.
- [B110] C. D. Giurcaneanu, I. Tabus, J. Rissanen (2000). MDL based digital signal segmentation. In *Proceedings of EUSIPCO-2000, X European Signal Processing Conference, Tampere, Finland, September 4–8*, vol. 1, pp. 339–342.
- [B111] J. Rissanen (2000). A generalized minmax bound for universal coding. In *Proceedings of IEEE International Symposium on Information Theory (ISIT2000), Sorrento, Italy, June 25–30, 2000*, IEEE Press, p. 324.
- [B112] J. Rissanen, B. Yu (2000). Coding and compression: a happy union of theory and practice. *Journal of the American Statistical Association*, 95, 986–988. Invited Year 2000 Commemorative Vignette on Engineering and Physical Sciences.
- [B113] Jorma Rissanen (2000). MDL denoising. *IEEE Transactions on Information Theory*, 46 (7), 2537–2543.
- [B114] Ioan Tabus, Gergely Korodi, Jorma Rissanen (2000). Text compression based on variable-to-fixed codes for Markov sources. In *2000 Data Compression Conference (DCC 2000), 28–30 March 2000, Snowbird, UT, USA*, pp. 133–142.

- [B115] Rissanen Jorma (2001). Information, complexity and the MDL principle. In *Cycles, Growth and Structural Change: Theories and empirical evidence* (Lionello F. Punzo, ed.), Routledge Siena Studies in Political Economy, Routledge, New York, NY, pp. 339–350.
- [B116] J. Rissanen (2001). Complexity and information in data. In *System Identification (SYSID 2000): A Proceedings volume from the 12th IFAC Symposium, Santa Barbara, California, USA, 21–23 June 2000* (R. Smith, ed.), vol. 1, Elsevier Science Ltd, pp. 1–6. Plenary paper.
- [B117] Jorma Rissanen (2001). Simplicity and statistical inference. In *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple* (Arnold Zellner, Hugo A. Keuzenkamp, Michael McAleer, eds.), Cambridge University Press, Cambridge, chapter 9, pp. 156–164.
- [B118] Jorma Rissanen (2001). Strong optimality of the normalized ML models as universal codes and information in data. *IEEE Transactions on Information Theory*, 47 (5), 1712–1717.
- [B119] J. Rissanen, B. Yu (2002). Coding and compression: a happy union of theory and practice. In *Statistics in the 21st Century* (Adrian E. Raftery, Martin A. Tanner, Martin T. Wells, eds.), *Monographs on Statistics and Applied Probability*, vol. 93, CRC Press, Boca Raton, FL, and American Statistical Association, Alexandria, VA, pp. 229–236.
- [B120] I. Tabus, J. Rissanen, J. Astola (2002). A classifier based on normalized maximum likelihood model for classes of boolean regression models. In *Proceedings of EUSIPCO 2002, XI European Signal Processing Conference, September 3–6, 2002, Toulouse, France*, vol. 1, pp. 119–122.
- [B121] I. Tabus, J. Rissanen, J. Astola (2002). Normalized maximum likelihood models for Boolean regression with application to prediction and classification in genomics. In *Computational And Statistical Approaches To Genomics* (Wei Zhang, Ilya Shmulevich, eds.), Kluwer Academic Publishers, pp. 173–196, 1st ed. Second Edition: Springer, 2006, pp. 235–258.
- [B122] Ioan Tabus, Jorma Rissanen (2002). Asymptotics of greedy algorithms for variable-to-fixed length coding of Markov sources. *IEEE Transactions on Information Theory*, 48 (7), 2022–2035.
- [B123] Petri Kontkanen, Wray Buntine, Petri Myllymäki, Jorma Rissanen, Henry Tirri (2003). Efficient computation of stochastic complexity. In *AI and statistics: Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, January 3–6, 2003, Key West, Florida* (Christopher M. Bishop, Brendan J. Frey, eds.), Society for Artificial Intelligence and Statistics, pp. 181–188.
- [B124] J. Rissanen (2003). Complexity and information in data. In *Entropy* (Andreas Greven, Gerhard Keller, Gerald Warnecke, eds.), Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, pp. 299–312.
- [B125] Jorma Rissanen (2003). Complexity of simple nonlogarithmic loss functions. *IEEE Transactions on Information Theory*, 49 (2), 476–484.
- [B126] B. Ryabko, J. Rissanen (2003). Fast adaptive arithmetic code for large alphabet sources with asymmetrical distributions. *IEEE Communications Letters*, 7 (1), 33–35.
- [B127] Ioan Tabus, Gergely Korodi, Jorma Rissanen (2003). DNA sequence compression using the normalized maximum likelihood model for discrete regression. In *2003 Data Compression Conference (DCC 2003), 25–27 March 2003, Snowbird, UT, USA*, IEEE Computer Society, pp. 253–262.
- [B128] Ioan Tabus, Jorma Rissanen, Jaakko Astola (2003). Classification and feature gene selection using the normalized maximum likelihood model for discrete regression. *Signal Processing*, 83 (4), 713–727. Special issue on genomic signal processing.
- [B129] Janne Ojanen, Timo Miettinen, Jukka Heikkonen, Jorma Rissanen (2004). Robust denoising of electrophoresis and mass spectrometry signals with minimum description length principle. *Federation of European Biochemical Societies Letters*, 570 (1–3), 107–113. Invited paper.
- [B130] J. Rissanen (2004). Complexity and information in modeling. In *Computability, Complexity and Constructivity in Economic Analysis* (K. Vela Velupillai, ed.), Blackwell Publishing, Oxford, chapter 4.

- [B131] J. Rissanen, I. Tabus (2004). Modelling with distortion. In *CISS 2004, 38th Annual Conference on Information Sciences and Systems, Princeton University, March 17–19*, pp. 560–563.
- [B132] Adriana Vasilache, Ioan Tăbuș, Jorma Rissanen (2004). Algorithms for constructing min-max partitions of the parameter space for MDL inference. In *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops SSPR 2004 and SPR 2004, Lisbon, Portugal, August 2004, Proceedings* (Ana Fred, Terry Caelli, Robert P. W. Duin, Aurélio Campilho, Dick de Ridder, eds.), *Lecture Notes in Computer Science*, vol. 3138, Springer, pp. 930–938.
- [B133] B. Barliga, I. Tabus, J. Rissanen, J. Astola (2005). Image denoising based on Kolmogorov structure function for a class of hierarchical image models. In *Proceedings of SPIE Optics and Photonics 2005: Algorithms, Architectures, and Devices and Mathematical Methods, Mathematical Methods in Pattern and Image Analysis, San Diego, California, USA, 31 July – 4 August 2005* (J. T. Astola, I. Tabus, J. Barrera, eds.), pp. 1–10.
- [B134] Petri Kontkanen, Petri Myllymäki, Wray Buntine, Jorma Rissanen, Henry Tirri (2005). An MDL framework for data clustering. In *Advances in Minimum Description Length: Theory and Applications* (Peter D. Grünwald, In Jae Myung, Mark A. Pitt, eds.), Neural Information Processing, MIT Press, Cambridge, Massachusetts, chapter 13, pp. 323–353.
- [B135] Gergely Korodi, Jorma Rissanen, Ioan Tăbuș (2005). Lossless data compression using optimal tree machines. In *2005 Data Compression Conference (DCC 2005), 29–31 March 2005, Snowbird, UT, USA*, IEEE Computer Society, pp. 348–357.
- [B136] J. Rissanen, I. Tabus (2005). Kolmogorov’s structure function in MDL theory and lossy data compression. In *Advances in Minimum Description Length: Theory and Applications* (Peter D. Grünwald, In Jae Myung, Mark A. Pitt, eds.), Neural Information Processing, MIT Press, Cambridge, Massachusetts, chapter 10, pp. 245–262.
- [B137] Jorma Rissanen (2005). An introduction to the MDL principle. In *Proceedings of the COBERA Workshop on Computable Economics, March 20–24, 2005, National University of Ireland, Galway, Ireland*.
- [B138] Ciprian Doru Giurcăneanu, Jorma Rissanen (2006). Estimation of AR and ARMA models by stochastic complexity. In *Time Series and Related Topics. In Memory of Ching-Zong Wei* (Tze Leung Lai Hwai-Chung Ho, Ching-Kang Ing, ed.), *Lecture notes—monograph series*, vol. 52, Institute of Mathematical Statistics, Beachwood, Ohio, USA, pp. 48–59.
- [B139] V. Kumar, J. Heikkonen, J. Rissanen, K. Kaski (2006). MDL denoising with histogram models. *IEEE Transactions on Signal Processing*, 54 (8), 2922–2928.
- [B140] Jorma Rissanen (2006). The structure function and distinguishable models of data. *The Computer Journal*, 49 (6), 657–664.
- [B141] I. Tabus, A. Hategan, C. Mircean, J. Rissanen, I. Shmulevich, W. Zhang, J. Astola (2006). Non-linear modeling of protein expressions in protein arrays. *IEEE Transactions on Signal Processing*, 54 (6), 2394–2407.
- [B142] Ioan Tabus, Jorma Rissanen, Jaakko Astola (2006). Nonlinear signal modeling and structure selection with applications to genomics. In *Advances in nonlinear signal and image processing* (Stephen Marshall, Giovanni L. Sicuranza, eds.), *EURASIP Book Series on Signal Processing and Communications*, vol. 6, Hindawi Publishing Corporation, New York, NY, chapter 4, pp. 79–102.
- [B143] Jaakko Astola, Gergely Korodi, Jorma Rissanen, Ioan Tabus (2007). DNA sequence compression – based on the normalized maximum likelihood model. *IEEE Signal Processing Magazine*, 24 (1), 47–53.
- [B144] Teemu Roos, Jorma Rissanen (2007). Conditional NML universal models. In *Information Theory and Applications Workshop (ITA-07), January 29 – February 2, 2007, University of California, San Diego*. Electronic publication. Available from <http://eprints.pascal-network.org/archive/00002972/01/CNMLITA.pdf>.

## Patents

- [C1] Glenn George Langdon Jr., Jorma J. Rissanen (1978), Method and means for arithmetic string coding. US Patent No. 4122440.
- [C2] Glen G. Langdon Jr., Jorma J. Rissanen (1981), Method and means for arithmetic coding utilizing a reduced number of operations. US Patent No. 4286256.
- [C3] Glen G. Langdon Jr., Jorma J. Rissanen (1984), High-speed arithmetic compression coding using concurrent value updating. US Patent No. 4467317.
- [C4] Glen G. Langdon Jr., Jorma J. Rissanen (1984), Method and means for carry-over control in the high order to low order pairwise combining of digits of a decodable set of relatively shifted finite number strings. US Patent No. 4463342.
- [C5] Glen G. Langdon Jr., Jorma J. Rissanen (1985), Adaptive source modeling for data file compression within bounded memory. US Patent No. 4494108.
- [C6] Kottappuram M. A. Mohiuddin, Jorma J. Rissanen (1987), Multiplication-free multi-alphabet arithmetic code. US Patent No. 4652856.
- [C7] Jorma J. Rissanen, Mati Wax (1988), Algorithm for constructing tree structured classifiers. US Patent No. 4719571.
- [C8] Glen G. Langdon Jr., Joan L. Mitchell, William B. Pennebaker, Jorma J. Rissanen (1990), Arithmetic coding encoder and decoder system. US Patent No. 4905297.
- [C9] Joan L. Mitchell, William B. Pennebaker, Jorma J. Rissanen (1991), Dynamic model selection during data compression. US Patent No. 5045852.
- [C10] Paul D. Healey, Jorma J. Rissanen (1994), Adaptive computation of symbol probabilities in n-ary strings. US Patent No. 5357250.
- [C11] Gilbert Furlan, Jorma J. Rissanen, Dafna Sheinvald (1996), Distributed coding and prediction by use of contexts. US Patent No. 5550540.
- [C12] Gilbert Furlan, Jorma Johannes Rissanen (1997), Quantization method for image data compression employing context modeling algorithm. US Patent No. 5640159.
- [C13] Gilbert Furlan, Jorma Johannes Rissanen, Dafna Sheinvald (1997), Distributed coding and prediction by use of contexts. US Patent No. 5652581.
- [C14] Rakesh Agrawal, Manish Mehta, Jorma Johannes Rissanen (1998), Data mining method and system for generating a decision tree classifier for data records based on a minimum description length (MDL) and presorting of records. US Patent No. 5787274.
- [C15] Jorma Rissanen (2006), Lossless data compression system. US Patent No. 7028042.

## **The 49 coauthors of Jorma Rissanen**

Ioan Tăbuș	17	Z. Vágó	2
Jaakko Astola	9	J. H. van Schuppen	2
Glen G. Langdon Jr.	7	B. Barliga	1
Bin Yu	5	Andrew Barron	1
László Gerencsér	4	F. Wyman Bostwick	1
Gergely Korodi	4	A. Chang	1
Edward J. Hannan	3	A. Hategan	1
Thomas Kailath	3	K. Kaski	1
Marcelo J. Weinberger	3	V. Kumar	1
Rakesh Agrawal	2	N. Mc Clamroch	1
Ronald Arps	2	Timo Miettinen	1
L. Barbosa	2	C. Mircean	1
Wray Buntine	2	Janne Ojanen	1
Peter E. Caines	2	C. Popeea	1
Meir Feder	2	Teemu Roos	1
Søren Forchhammer	2	B. Ryabko	1
Ciprian Doru Giurcăneanu	2	G. Shedler	1
Jukka Heikkonen	2	I. Shmulevich	1
Petri Kontkanen	2	Terry Speed	1
Lennart Ljung	2	Stephen Todd	1
Manish Mehta	2	Adriana Vasilache	1
K. M. Mohiuddin	2	Mati Wax	1
Petri Myllymäki	2	V. Wertz	1
Eric Sven Ristad	2	W. Zhang	1
Henry Tirri	2		