Mathematical Conversations

Sergio Verdú: Wireless Communications, at the Shannon Limit >>>

Sergio Verdú

Interview of Sergio Verdú by Y.K. Leong (matlyk@nus.edu.sg)

> *In the beginner's mind there are many possibilities, but in the expert's mind there are few.*
> – Shunryu Suzuki (1904 – 1971), Japanese Zen Master

Sergio Verdú is world-renown for pioneering the field of multiuser detection in wireless communications and for fundamental work on data transmission and compression in information theory.

His theoretical doctoral research has a tremendous impact on communications technology with numerous applications in mobile cellular systems, fixed wireless access, high-speed data transmission, satellite communication, digital television and multitrack magnetic recording. His book *Multiuser Detection* published in 1998 is now a modern classic. His research papers have received many awards from scientific and professional bodies. He has also received several awards for professional education and outstanding teaching. The prizes, awards and accolades bestowed on him are indeed too numerous to list; the latest in 2007: election to the US National Academy of Engineering and the Claude E. Shannon Award, the highest honor in information theory.

On the faculty of Princeton University since 1984, Verdú is Professor in the Department of Electrical Engineering since 1993. He is also a core faculty member of the Program in Applied and Computational Mathematics. He is known for his personal zeal in advisory and organizational work (not only in the United States, but from South America across Europe to Asia) in advancing and promoting science and technology. He has served as President of IEEE Information Society and he serves on the editorial boards of leading journals in his field, in particular, for IEEE. His scholarship and scientific charisma have led to many visiting appointments and invited lectures around the world.

Verdú was an invited speaker at the Institute's program *Random Matrix Theory and its Applications to Statistics and Wireless Communications* (26 February – 31 March 2006). He was interviewed on 26 February 2006 by Y.K. Leong on behalf of *Imprints*. The following is an edited and enhanced version of the interview in which he traced his scientific path from humble beginnings in Barcelona, Spain to prominence in the world's leading centers of communications research in the United States. It resembles a classic Spanish narrative that spans a wide spectrum from human passion to intellectual vision set on a scientific stage for exploring the physical possibilities of communications at the edges of the theoretical limits of information theory.

Imprints: In your undergraduate training in Spain, you had already specialized in telecommunication engineering. Why did you choose this particular branch of engineering?

Sergio Verdú: I decided to become a telecommunications engineer when I was 7 years old. My father gave me a toy – a kit with which you could build radios and all sorts of electrical devices – and I was hooked. My father was very good with electrical gadgets. As a child I was always immersed in electronics. At that point I decided to be an electronics engineer and I never wavered from that.

I: Was your father an engineer?

V: No, my father had very little formal education. His childhood was spent during the Spanish Civil War. He suffered a lot and went through tragic circumstances. He was a self-made man, a very fine man. He really had a lot of influence on me even though he died in an automobile accident when I was 11 years old.

I: What attracted you to go to the United States (in particular, the University of Illinois at Urbana-Champaign) for your graduate studies?

V: Going back to my early youth, I guess that's where you would find the traces for all these decisions. My parents decided that I would get an English tutor when I was 6 or 7 years old. From then on until when I was in high school, I had an English tutor: a Spaniard, she was not a native English speaker and I guess that accounts for my less than perfect accent. It gave me an edge over everybody else who was just learning English in school. I was fascinated with things American and in particular with the space program. When

Continued on page 9

Continued from page 8

I was 14, I jumped at the opportunity to spend a summer in Rockville Center, Long Island in 1973. It was an excellent opportunity to see a world completely different from the backward country where I had grown up: Spain was under a fascist dictatorship from 1939 to 1975. At that time things like color television were completely new to me. I remember being fascinated by the Watergate affair that was going on at that time. The fact that a country could be so open politically while undergoing a painful episode and still able to do it with a sense of humor was a revelation. Although I tried, practically the only American activity that to this day I never could really get interested in was baseball. At that time it was certainly not very common for Spanish students to do their graduate studies in USA. In fact, I didn't know anybody who had done that. I ended up at the University of Illinois at Urbana-Champaign partly because as an undergraduate I had worked a lot on computer-aided design. I was very much into the design and analysis of electrical circuits using the computer. A lot of prominent people in that field have been at Urbana-Champaign. I was also admitted at Stanford and, of course, I knew some of the professors there but perhaps not as much as the ones at Urbana. But at that time, I already knew that I had done enough programming and hacking in my life. I really wanted to do theoretical work and I really wanted to do communications theory and information theory. Of course, as an undergraduate I had already heard of Shannon, and one day when I was discussing my options to go to graduate school in the US, one of my professors said, "Well, you know, Claude Shannon was at the University of Illinois." I said, "Oh, okay." That clinched the decision for me.

I arrived in the United States in1980 during the Presidential campaign between Ronald Reagan and Jimmy Carter, and Urbana-Champaign was pretty shocking to me. It was so unlike the atmosphere in the New York area that I had seen in `73 and, needless to say, very different from the big European city life that I had been exposed to. In addition to the geographical isolation, the religious atmosphere of the place was really striking. When I got there, I asked, "So when did Claude Shannon teach here?" and nobody knew about Claude Shannon having been there. One day, browsing in the university bookstore, I picked up a copy of Shannon's *The Mathematical Theory of Communication*. It had been reprinted by the University of Illinois Press.

I: Then you went to Princeton immediately after Illinois?

V: Yes, the day after I defended my PhD thesis. My wife Mercedes and I drove our Chevy to New Jersey.

I: You didn't go back to Spain?

V: No. I always wanted to remain in the United States. I had a Fulbright Fellowship. That gave me a lot of trouble because Spain, concerned about the brain drain, refused to give me permission to stay in the United States. But, after a long, complicated process through the State Department and the Department of Justice involving senators and so on, I was granted a waiver of the requirement to return.

I: Your doctoral research pioneered the field of multiuser detection. Could you tell us something about it? Were you excited and surprised by your work at that time?

V: Yes. At that time (in the early 80s), I had worked for my masters' thesis in minimax robustness. This was a field that originally started in statistics with the work by Huber in the 70s. Then there was a lot of work in engineering (particularly by my advisor Vincent Poor) applying Huber's theory to robust estimation, robust detection and so on. Vincent Poor mentioned that in spread-spectrum communications, they were modeling the multiaccess interference as white Gaussian noise, and although this seemed to be a pretty good modeling assumption, perhaps there was some room to apply robust statistical methods to account for the deviation from the central limit theorem. I started to look at it from that angle, but then I quickly realized that that was not the right approach and that a completely new approach had to be taken. Then I obtained the optimum multiuser detector, and that became the beginning of my PhD thesis. The interesting thing was not only the structure of the receiver but the fact that in many cases you could achieve single-user performance. The gain was remarkable and much more than what we expected. That was the beginning of multiuser detection. At that time, nobody was paying any attention to it. Spread-spectrum research was pretty much dominated by military funding and did not have the vibrancy it acquired later on, thanks to the ascent of wireless telecommunications and CDMA wireless commercialized by Qualcomm.

I: Your doctoral work was not classified?

V: The university would not allow any classified research. It was actually good for me that at the beginning it did not attract any interest. It was only years later that multiuser detection became a very vibrant research field with a lot of research citations to the early work I had done in the early 80s. In 1998, I published a book, essentially a compilation of my work and my teaching of the subject. But sometime in the late 1980s it ceased to be my primary research focus. Perhaps if the success had been immediate, then I would have devoted a lot of my efforts into that and less into information theory, which eventually became my primary field of interest. I think it was propitious, and interestingly, the time constant from inception of ideas to implementation of these ideas in that particular field was very, very long. It's only recently that there has been motivation and success

Continued on page 10

Continued from page 9

in industry implementing multiuser detection. One of the drivers has been multi-antenna systems where there is interference between signals transmitted by different antennas. Qualcomm, the proponent of CDMA cellular wireless, came up with a second-generation cellular wireless with rather old signal processing algorithms. It didn't use any multiuser detection, but they have announced recently that they are using these methods in their third-generation chips. These are channels where bandwidth and power are resources to be conserved. And one of the lessons that Shannon taught us is that you have to exploit the fine details in your model (in this case multiuser interference) to squeeze the most out of the channel resources.

I: It is now commonplace?

V: It depends on which area. Although the systems were not designed with the idea that you would have sophisticated receivers taking into account multiuser interference, both in third generation CDMA and in Digital Subscriber Loops (high speed data through telephone copper wires), they are starting to implement it. Multiuser detection is commonplace in the multi-antenna receivers where you can get substantial gains in capacity taking into account interference proceeding from different antennas. There are also chips that take into account intertrack interference in magnetic recording. It is always just a matter of time until the maturity of technology puts a stop to the waste of bandwidth/power.

I: Have you ever considered working in industry?

V: No, I was always very much an academic type. I like the freedom to pursue my own ideas and my own work. I also like to interact with young people. Not having a boss is nice too.

I: You seem to be equally comfortable with mathematics and engineering. How do you manage to reconcile their two different approaches to problem-solving – approaches that are apparently poles apart?

V: Strangely enough, they are not very different because the way you approach problems is essentially the same in both fields: going back to the basics. As much as I can, I always try to avoid carrying a bag of tricks that I can apply from one problem to another. I have actually moved quite a bit from problem to problem, and there's a lot of pleasure starting on a problem from scratch that I really didn't know anything about. Like the Zen philosophy says, in the mind of the beginner the possibilities are endless. A lot of important contributions are made by people who have just entered the field. Learning new mathematics is a delightful reward. Technology points out what next to learn; for example, my work on random matrices – which is why I am here now – was motivated by wireless communication systems. The type of research that excites me is mathematically challenging and relevant to the real world. Claude Shannon was the archetypical seamless combination of mathematician and engineer.

I: Do you think that, in general, engineers have as much mathematical training as they should have?

V: Mathematical training is like wealth, nobody has enough of it. The thing about this discipline that we call electrical engineering is that its unifying theme (electricity) goes back to the 19th century and is now completely obsolete. But our engineering training gives you a lot of versatility to deal with very different problems. To give you an example, two of my graduate students are finishing their PhDs in information theory this summer and are joining Goldman Sachs and Credit Suisse. Electrical engineering undergraduates may not get as much mathematical training as they would need to be professors doing research on say telecommunications. That mathematical training they will have to get later on in graduate courses and on their own. But electrical engineering undergraduates do get very strong training in problem-solving, and this gives them a lot of options.

I: It seems that one perception about the mathematical training for engineers is that they are more interested in sort of recipes or a bag of tricks for solving problems.

V: The training in engineering is very different around the world. Some of the European systems tend to have a kind of dichotomy. In the first two years of engineering, they are very mathematically oriented, and then later the subjects become very practically oriented. For example, I had to take two semesters of television – something that would be completely unheard of in the US. I think that in the US, perhaps because the professors are much more research oriented than in other places, we tend to be more mathematically oriented, at least those of us on the applied mathematics side of electrical engineering, like communications, control and signal processing.

I: Do you work directly with hardware engineers to create the technology?

V: No, not really. By the way, the dichotomy between hardware and software is fading. It is always important to be aware at any given time what the technology can deliver so that you know whether the solutions you are coming up with are solutions that can be implemented now or in 20 years' time or perhaps the technology in a certain field has progressed so much that you can implement things that are much more sophisticated than what people are

Continued on page 11

Continued from page 10

implementing right now. So it's very important to have a sense of what technology can deliver even if normally we don't collaborate in research with people working in hardware.

I: What is the "biggest" unsolved theoretical problem in communications technology?

V: The biggest success story of Shannon's theory has been in point-to-point communications. Shannon's theory has been instrumental in anything that has to do with modems, wireless communications, multi-antenna and so on. But network information theory has proved to be a particularly tough challenge. Shannon was the first to formulate the problem, or at least the building blocks, in 1961. Instead of having one transmitter and one receiver, you have a bunch of transmitters and a bunch of receivers, and you may also have some nodes in between that act as relays, and some of those nodes may also be sources or sinks of information. You could think of a very general topology and you would like to know what are the best rates of information, what are the distinguishable signals that you can send. This is something that we still don't know.

Another important technological challenge is data compression of audio and video signals, which in my view is still in its prehistory. Even though Shannon also gave the fundamental principles of this discipline, information theory has not had nearly as much impact as it has had in channel transmission or in text/data compression. I think the reason is that we do not yet have a good understanding of human vision and hearing, and even the little we know is hard to marry with the available theory.

I: The point-to-point problem is solved?

V: We understand it a lot better. Shannon gave us the point-to-point framework, but he didn't give us all the solutions. Finding the capacity of a particular point-to-point communications channel may be extremely challenging and, in fact, the capacity of some very simple channels is still unknown.

I: Do you agree that engineers are very focused in their research in the sense that they try to solve only problems that are of immediate practical concern in contrast to physicists who try to answer fundamental questions that are not immediately applicable?

V: No, I don't agree. Shannon was the primal example of an engineer who would explode this myth. Many of us who are working in theory are accused, more often than not, of doing exactly the opposite: of solving problems that are of no immediate practical concern and that may become relevant only in the distant future or never. Those of us who have followed in Shannon's footsteps have an appreciation for beauty and elegance and for the fact that beautiful and elegant results sooner or later become practical. So you need to have some faith even though what you are working on now is not of immediate practical concern. You may be interested in it not because of some technology out there clamoring for solution, but because of its beauty.

I: It appears that you are a mathematician first and then an engineer.

V: I would say first an engineer, then a mathematician and then an engineer. I have come full circle. My doctoral thesis had an important component in developing algorithms, and also a lot of analysis but I had this nagging feeling that it was not mathematical enough for my taste. When I got into information theory I became quite theorem-proving minded. But, the thrill of coming up with new algorithms is something I have come to appreciate later, more so in recent years. When I was younger, I had the idea that if I cannot prove a theorem about something, then I don't want to do research on it. Now my outlook has evolved. Of course, I still like to prove theorems, but I have also done some recent work that is algorithmic and I enjoyed it very much.

I: Can you tell us something about your present research interests and the problems you are working on?

V: People say that the interesting problems are at the boundary between disciplines. This is actually true sometimes. One of my current interests is the boundary between information theory and estimation theory. A couple of years ago, we found a very basic formula that connects some basic quantities from information theory and estimation theory. Capitalizing on this formula, we gave some simple proofs of a probability theory result on the monotonicity of the non-Gaussianness of the sum of independent random variables as well as a famous result from Shannon's 1948 paper, called the entropy-power inequality. We also came up with a new universal formula in continuous-time nonlinear filtering, as well as an algorithm to minimize transmitted power. All those come from this innocent-looking formula. As usual, there is nothing more insightful and practical than a pretty formula.

Random matrix theory has been very rewarding. I got into random matrix theory around 1997. When I was finishing my book, I was fortunate to become acquainted with Marchenko-Pastur's theorem and I included it in Chapter 2. Since then there has been an enormous interest and excitement. It is challenging to get into this theory. Even though its early history developed in the 50s and 60s, the core results are quite recent. It's only in the last 10 years or so

Continued on page 12

Continued from page 11

that there has been a lot of interest in it from contemporary mathematicians.

I: Is random matrix theory applicable in engineering?

V: Very much so. It's applicable and fundamental in wireless communications. The first application was in the capacity of multiple antenna systems. In Bell Labs, Foschini and Telatar realized that in the presence of electromagnetic scattering when you have multiple antennas the channel capacity can be much larger than if you have single-antenna transmitter and receiver. Random matrix theory is fundamental in this realization, and also in the analysis of the fundamental limits of spread-spectrum in wireless communications.

I: Have there been any breakthroughs in random matrix theory?

V: The great breakthrough, at least for applications in wireless communications, was in 1967 in the work of Marchenko and Pastur in the Soviet Union. That was an amazing piece of work. It was completely unknown for many years. In 1986, I looked at a random matrix problem that I wanted to solve. I looked in the literature (of course, that was before Google) and I could find nothing. The work on random matrices that I could find was completely orthogonal to what we needed. Physicists and mathematicians were rediscovering the Marchenko-Pastur result in the late 80s and 90s. Lately there has been a lot of excitement in a new mathematical field called "free probability" and one of its main applications is in random matrix theory. Wireless communications and information theory have been one of the main propellers of work in this theory. We are not just consumers of this kind of result; we have also been able to pose new questions and solve some of these problems.

I: So, in a sense, wireless communications has affected the development of random matrix theory.

V: Oh, yes, for sure. You see this pendulum of interaction in other fields. Information theory was very much influenced by ergodic theory, and also the other way around. Kolmogorov made a fundamental discovery in ergodic theory thanks to information theory.

I: Do you do much consultation work for industry?

V: I occasionally have done work with people in research labs such as Bell Labs, Hewlett Packard, and Flarion, which was recently acquired by Qualcomm. When I was doing work in Hewlett Packard, the group there was very theoretically inclined. I was actually kind of like the guy who was pushing for us to do more algorithmic work rather than theorem-proving. It was particularly rewarding to be associated with Flarion because you see the thrill of seeing brilliant ideas being implemented in a very short period of time.

I: Do you have any patents?

V: Traditionally, being more academically oriented towards peer publication, I have not pursued patents at Princeton. But yes, I do have quite a few patents granted or pending both through Bell Labs and Hewlett Packard.

I: Do you think that technology will be able to catch up with the theoretical advances in science and technology or even mathematics?

V: Well, Shannon's theory is a good example of a theory that at the beginning created a lot of enthusiasm. Shannon became an instant celebrity. Then, for a few years, people were asking the question, "If this is so good, how come it hasn't seen the light?" Of course, what happens is that it came well before its time, well before technology was ripe to be implemented. It took a long, long time for implementable codes to achieve Shannon's limits. In data compression they did not appear till the 70s and in data transmission until the 90s.

I: When was Shannon's theory put forward?

V: 1948, so it took a long time. That's a powerful lesson because everybody knew that these were very powerful ideas. For decades, there was a lot of unsuccessful work in trying to design codes that would approach Shannon's limits. When there are theoretical breakthroughs and when we are able to solve problems of a fundamental nature, then just because technology doesn't seem to be on the near horizon to be able to implement those ideas or formulas, it doesn't mean we should give up and say, "Okay, this is a dead field because we have given it long enough time and technology has not implemented it, and therefore it is hopeless." I think information theory is a great lesson in that respect. In communications, we have a limited piece of spectrum that we can only use with given resources, and there is an enormous economic incentive to use that spectrum as efficiently as possible. So when you have a theory like information theory that sets fundamental limits, there is an enormous incentive to get as close as you can to those limits.

I: How would one make the theoretical work drive the technology faster? Is that possible?

V: Well, it is possible. In certain areas, it has been enormously successful, for example, in modems, in work that was published in the *Information Theory Transactions*. Four

Continued on page 13

Continued from page 12

years later, you could buy modems that were implementing those ideas for a hundred dollars. That is a field where the time constant is much faster. In fields like cellular wireless, the technology transfer has been a lot slower. Developed in the late 80s, second-generation wireless systems were predicated on technology that was really old (a lot of it 50s, 60s). A revolution happened in the 1990s with the advent of a class of channel codes called the turbo codes. In the beginning, they were not very appealing to the theoreticians because these codes came very close to the Shannon limit, but nobody could explain why. We couldn't come up with theorems that would say. "Hey, of course, this is why they do work." Now we understand them better. Actually, they vindicate Shannon because he came up with a theory for what the best code could do without the benefit of knowing a single code except possibly for the simple Hamming code that was developed at the same time. He said, "Well, I don't know how to construct the optimum code but I can show that a construction where the codes are chosen blindly at random, performs close to optimum on the average." The problem is that if you choose a code at random, it cannot be implemented because it doesn't have structure. So these new codes that go back to the 1990s turn out to have enough structure that you can implement and encode them in linear time and at the same time they have enough randomness like Shannon originally said in 1948 to be close to the best.

I: So without those codes the cellular revolution would have been impossible?

V: The first digital systems used codes that were really far from capacity. Early in the game in the design of codes, people took a turn away from Shannon's random codes. Coding theory became a geometric discipline, very combinatorial, not so probabilistic. Now we are going back to the roots. Those geometric constructions that emphasize minimum distance properties of codes are not the ones that achieve capacity. They are very interesting mathematically but are not the ones that turn out to come closest to Shannon's fundamental limits. With the new codes you can increase the efficiency quite a bit. What is surprising is that there was nothing inherently there to prevent people in the 1960s to come up with these codes. Actually, Gallager at MIT had come up with random-like codes in the early 60s but he abandoned them because they thought they could never be implemented and that they were too complicated. They are actually not too difficult to implement. The key is not to attempt optimum decoding because that is too expensive. With a judicious choice of code, the life of the decoder is a lot easier, and near optimal decoding is feasible. The bottom line is that in linear time you can come very close to the Shannon limit. People are now using cellular phones that incorporate these codes.

I: Is the Shannon limit a real physical Heisenberg-type limit or is it a Gödel-type logical limit?

V: The short answer is: information theory is a chapter of probability theory, which in turn is a chapter in mathematics. The starting point is a stochastic model for the information source and a stochastic model for the channel. Are those models relevant to the real world? If they weren't, your cellphone would not work. Having said that, since 1948 there have been enormous strides in information theory dealing with uncertainty in nonprobabilistic ways. An example is the theory of algorithmic complexity which is devoid of any probability, and was put forward by Kolmogorov, the father of modern probability theory.

I: Do you think that a revolution in wireless communications would follow in the wake of breakthroughs in nanotechnology?

V: The radiofrequency spectrum usable in wireless communications is rather limited. Information theory tells us the fundamental capacity of the medium. We cannot go beyond it no matter how fast the computing technology. But let me address the question from a broader perspective: why can a DVD contain a lot more music than a CD? The compression technology of the CD dates back to the 1930s. By the time the DVD was developed 15 years after the CD, lossy compression was much better understood, and the optical recording devices were also quite a bit more advanced. So the engineer reaps benefits from both applied physics and applied mathematics. For the information theorist, new physical devices mean new communication channel models, with a capacity to be discovered. So I think information theorists are going to be around for a long time.