S. Caban, J. A. Garcia Naya and M. Rupp, "Measuring the physical layer performance of wireless communication systems: Part 33 in a series of tutorials on instrumentation and measurement," in *IEEE Instrumentation & Measurement Magazine*, vol. 14, no. 5, pp. 8-17, October 2011, doi: 10.1109/MIM.2011.6041377.

Measuring the Physical Layer Performance of Wireless Communication Systems

Sebastian Caban, José Antonio García Naya, and Markus Rupp

Part 34 in a series of tutorials on instrumentation and measurement

Measuring the physical layer performance of wireless radio communication systems is one important step to clearly understand their behavior in real-world environments, that is, non-simulation environments. Unfortunately, measurements in wireless communications are extremely expensive and time-consuming. In fact, they require researchers to deal with non-artificial, realizable, real-world problems. They also require these researchers to have experience in computer engineering, telecommunication engineering, electrical engineering and often even mechanical engineering. Finally, they require teamwork to set up complete systems instead of dealing with single, isolated, numerical environments.

In this era of low-cost and ever more powerful personal computers, more and more research groups are abandoning hardware-based research in favor of "simpler," simulation-based research. If this tendency continues, soon, only a few research groups will be left that publish results based on large-scale measurement setups. One might think that measurements have lost most of their appeal.

At the end of the day, each research group has to decide if it is up to the challenge of measurements that costs time and money. To ease the non-financial part of this challenge, we offer some advice and a few pointers based on our experience in this tutorial [1].

Problem Statement

Let us examine the wireless communication system in Fig.1. This simple set-up can be extended easily to resemble a complete wireless communication system, and at the same time, it also serves as an example for the tutorial.

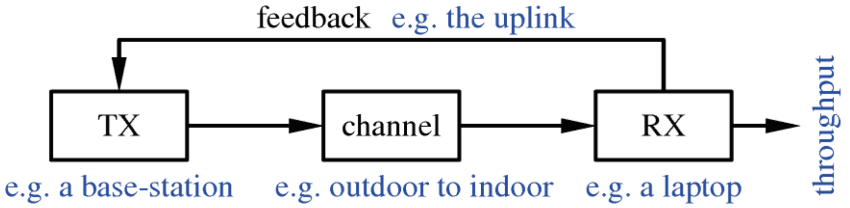

Fig. 1 A sample problem set-up; the downlink of the physical layer of a wireless communication system

The set-up consists of three blocks. Data are generated in the block named 'TX'. These data are next transmitted over a wireless radio channel. Finally, the data received are processed in the ‘RX’ block to calculate the figure of merit of the data transmission, namely the throughput typically expressed in Mbit/s. We are interested in the physical-layer throughput that can be achieved in a given, specific, real-world scenario – with actual physical channels, transmitters, receivers, and a standard-conform transmission.

Measurement Results

Fig. 2 shows measurement results obtained in 2009 in downtown Vienna, Austria. It compares the average throughput of a “2x2-transmission” (that is, a transmission that uses two antennas at the transmitter and two antennas at the receiver site) with the reference of a “1x2 transmission” (that uses only one antenna at the transmitter but still uses two antennas at the receiver site). The dashed curve represents the throughput difference, the gain in throughput obtained by one additional antenna at the transmitter, between these two transmissions. The dotted curve represents the unconstrained channel capacity of the 2x2 transmission, that is, the maximum theoretical throughput allowed by the channel for the 2x2 transmission under investigation.

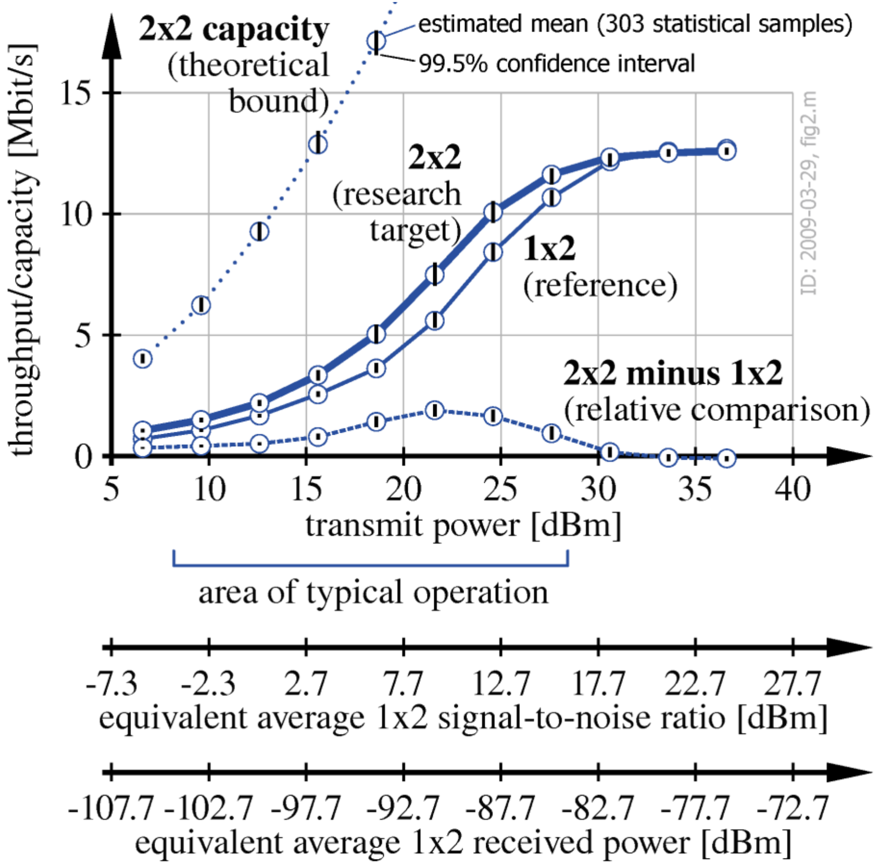

Fig. 2. A sample measurement result.

As simple as it may look at first glance, it took us several years to obtain such figures efficiently in a reproducible and repeatable fashion for modern wireless communication standards such as WiMAX, HSDPA (the example in Fig. 2), LTE, and LTE advanced. As space is limited, we are only able to roughly sketch what steps we took to obtain such measurement results in real-world scenarios.

Still the First Step: Simulation

Starting with the work of Ronald A. Fisher in 1935, there exist many excellent books on the design of experiments [2], [3]. They all advise us of the importance of carefully planning experiments and of following the scientific method. Reality, however, is different. Expensive hardware is often purchased only to lie unused in the corner of a laboratory. Therefore, browsing catalogs and homepages of several manufacturers to buy “fancy hardware” is not the most intelligent starting point for experimentally evaluating wireless communication systems. Rather, after a thorough experiment design, simulation is the first and most important step to avoid unnecessary costs as well as efforts. Such simulations should not only concentrate on the “kernel” of the problem to be investigated but should be designed to simulate the measurement environment.

Let us now reexamine the set-up shown in Fig. 1. Being only one among many possible setups, we use it as an example to highlight important properties of the simulation to be carried out (see Fig. 3).

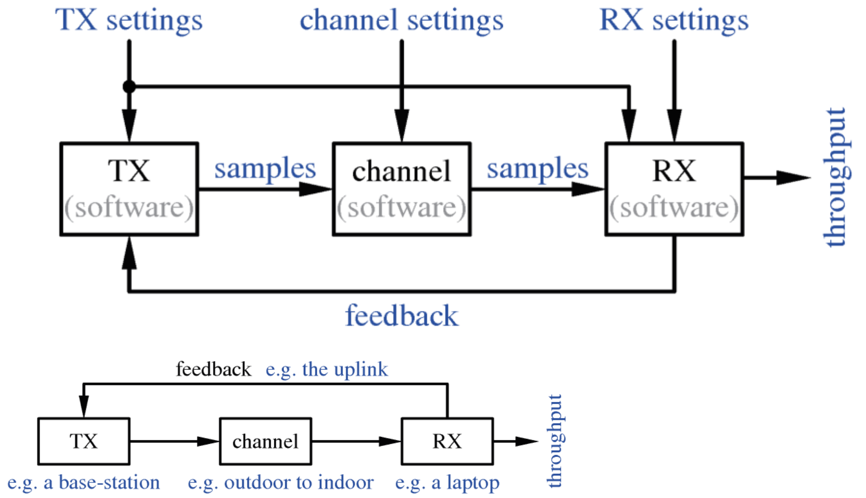

Fig. 3. A sample simulation/measurement set-up. Every block represents a software-function/hardware-block. Note the limited information passed between the different blocks.

At first, the whole set-up has to be split up into distinct blocks. These blocks should be implemented as separated software modules, for example functions. A strict separation of the different blocks is necessary, as these simulated blocks will be mapped in a next step onto different hardware devices such as a transmitter and a receiver. These devices are often physically separated; therefore, global information cannot be used to control them. Only for settings that do not change during the course of an experiment (for example, the name of a computer or the identifier of the measurement), global information may be used to simplify the measurement software. However, because it is fixed throughout an experiment, such global information decreases flexibility and should be avoided as much as possible. Instead, data structures containing the settings should be passed between the different blocks. This leads to increased flexibility when very different sets of settings have to be selected repeatedly during a measurement (for example, the set of transmit power levels, or the set of antenna configurations).

Often, the information passed between different blocks has to be very limited, as it will be the case in the final experiment. There, signaling bandwidths may be limited and subject to significant latency. Sometimes, it may also happen that desired information cannot be passed at all between different blocks as this information will not be available in the measurement. For example, in Fig. 3, channel properties are generated inside the channel without the possibility of passing them to and from the channel. In the simulation, as in the measurement, the true noise power and channel are both unknown. The noise realization is also not kept equal for consecutive blocks as preferred in Monte Carlo simulations that compare different transmission schemes. On the other hand, the same set of blocks (chosen from a large pre-generated set of blocks) may always be transmitted to avoid the spending the time to generate them on-line at the transmitter.

Still, even if a simulation is carried out with the greatest care, unforeseen effects may be revealed later in the measurement. For example, using commercial-grade radio-frequency components may result in in-phase to quadrature imbalance and non-flat frequency responses next to nonlinearities. Whether such effects should be calibrated out of the hardware, if this is possible at all, depends on the problem statement. Whether such effects should be included in or neglected in the simulation once the effects are discovered depends on the degree of their impact on the result.

Other than this, one may be able to infer key parameters of the hardware set-up from the simulation. Accurate knowledge of such parameters, for example, the peak signal-to-noise ratio at the receiver, may guide the design and implementation of the hardware required for the measurement. Still, in academic research, it may sometimes be better to design the hardware one magnitude more precise than required such that a “bound” can be measured and compared to a practical, meaningful implementation.

Last but not least, one should not forget the unloved step of verification and validation that should be natural in research.

How to Combat Fading?

Isolated, absolute throughput measurements of wireless communication links tend to lead to different results in practice depending on when and where exactly they are carried out. This happens for two reasons:

- First, channels observed in wireless communications experience “small scale fading”. In other words, the received signal power may change by several orders of magnitude when moving the receiver or objects in its vicinity only slightly. For a signal transmitted at 2.5 GHz, moving objects on the order of 6 cm, namely half a wavelength, can cause a large change in the power of the received signal.
- Second, channels observed in wireless communication experience “large scale fading”. That is, measuring in different scenarios, for example different buildings or streets, will lead to fundamentally different received signal strengths depending on the observed shadowing.

To combat large and small scale fading, one may carry out so-called drive test measurements. That is, one can make measurements at several positions on a path driven with the receiver or the transmitter and use the data recorded to plot empirical cumulative distributions of the values observed.

In this tutorial, we focus on another way of dealing with fading in wireless measurements. The basic idea is simple, and Fig. 2 shows the results of such a measurement taken using the technique we describe now. Detailed reasons for the steps we recommend taking will be given in the following sections.

- Step 1: Fix the physical position of the transmitter TX (the base-station).
- Step 2: Fix the total transmit power to a specific value, that is, the total power transmitted on all available transmit antennas expressed in dBm. Using a power meter, this power value can be easily measured at the input of the transmit antennas and plotted on the abscissa of Fig. 2.
- Step 3: Move the receiver (the laptop) to a distinct position within an area of approximately three times three wavelengths (36 cm times 36 cm for a transmission at 2.5GHz) by the use of an accurate XY-positioning table. By doing so, the receiver experiences small-scale fading but avoids large-scale fading.
- Step 4: Transmit all schemes of interest. In the example, to measure the values plotted in Fig. 2, we conducted a “2x2 transmission” and a “1x2 transmission” using exactly the same set-up: exactly the same transmitter position, total transmit-signal power, and receiver position. But more important, as we transmit all schemes consecutively within the channel coherence time, they experience exactly the same channel realization.

- Step 5: Repeat Steps 3 and 4 to obtain more throughput values of the same small-scale fading scenario at the same transmit power. Each of these throughput values represents a statistical sample and, if not too many of them are taken in the limited area available, the samples are statistically independent [4].

  The results shown in Fig. 2 were obtained from 303 statistical samples taken uniformly in a grid of three times three wavelengths. Choosing random positions would also have been an option, but when measuring many realizations, systematic sampling tends to minimize the correlation between the samples obtained. In the end, the choice of sampling procedure is a tradeoff between the loss of precision due to correlation and possible errors introduced by a systematic sampling approach [3, page 221].
- Step 6: Average the (in our example 303) throughput values for each scheme to obtain the best estimate for the "mean throughput" at each transmit-power level (the circles shown in Fig. 2), as no additional information other than independent samples is available. Of course, any other estimator, as for example "the median", would also work with this methodology.
- Step 7: Bootstrap the throughput values to obtain $\mathrm{BC_a}$ confidence intervals for the mean (the black vertical lines shown in Fig. 2) [5], [6].
- Step 8: Repeat Steps 2 to 7 for different transmit power levels (the abscissa in Fig. 2) to obtain the mean throughput over transmit power. In the measurements we performed, we achieved these different power levels by precisely attenuating the transmit signal using a combination of digital and analog attenuation which we have calibrated for linearity prior to the measurement. Note that this step accounts for the avoided large scale fading.
- Step 9: Name each measurement uniquely (see the measurement identifier "ID: 2009-03-29" at the right-hand side of Fig. 2). Next, automatically backup all scripts and source codes with the measurement results, and take many pictures of the measurement set-up. Experience has shown that when measurement results are reused years later, such information is indispensible, especially when details of the measurements become relevant that were not important during the initial measurement, as for example, if a door was open.

Over the course of the last few years we have repeatedly and successfully tested this methodology on most modern wireless communication systems in several urban and non-urban scenarios. In contrast to other methodologies, for example drive test measurements, this methodology does not allow for an *efficient* comparison on how a transmission scheme behaves in very different scenarios, even though this would be possible by repeating all of the steps above over and over. On the other hand, if such a comparison is not intended and only the behavior of a communications system in specific scenarios is of interest, the above-presented methodology may be the methodology of choice.

Why to Measure Over Transmit Power?

In wireless communications, the figure of merit (for example, the throughput) is usually thought of as a function of signal-to-noise ratio or signal-to-interference noise ratio (the additional abscissas in Fig. 2). We do things a bit differently for a few reasons.

First, modern wireless communication standards employ multiple antennas at the transmitter, and combine this change with techniques such as channel adaptive precoding or beam-forming. While keeping the transmit power fixed, such techniques adaptively modify the transmit signal to increase the signal-to-noise ratio at the position of the receiver. As a consequence, the performance of the wireless transmission is improved. Plotting the corresponding performance curve over the signal-to-noise ratio at the receiver would cancel out all of this improvement, thus giving a misleading picture of reality (although still a correct picture for a given signal-to-noise ratio).

Second, assume that your task is to compare two equal wireless communication links employing different antennas at the transmitter. The first antenna has a gain of 20 dBi, while the second one offers a gain of 5 dBi. Consequently, the communication system with the 20 dBi antenna will experience a higher throughput. But, if one plots the throughput of these two communication systems over the signal-to-noise ratio at the receiver, their performance will be exactly equal (as the scenario is linear and equal in both cases, the throughput is only a function of the signal-to-noise ratio). This reasoning also holds for different antenna orientations, different antenna polarizations, different noise figures of the receiver, and for different shadowing by walls, for example.

Third, channels in wireless communication experience fading. This leads to an instantaneous signal-to-noise ratio at the receiver that is different for measurements at different receive antenna positions. Therefore, in simulations, researchers usually plot the ‘expected signal-to-noise ratio’ on the abscissa which may typically even be known beforehand as the channel follows a given known model with clearly defined parameters. In measurements, however, this value first has to be estimated over time and space, assuming that the system observed is ergodic. But even more problematic, a specific value for the expected signal-to-noise ratio at the receiver cannot be set but only be observed with the precision of possibly only a few measurements.

On the other hand, even if we measure over transmit power, we still express the results additionally in terms of signal-to-noise ratio at the receiver. Therefore, we plot the signal-to-noise ratio (SNR) in Fig. 2 only for the 1x2-link and call it “equivalent average 1x2-SNR”. As the 2x2 link operates using channel adaptive beamforming, its signal-to-noise ratio at the receiver is significantly better, and, therefore, the throughput curve is shifted to the left. This effect is not cancelled out by plotting over transmit power, rather than plotting over signal-to-noise ratio.

Everything is Relative

Let us rethink the setup in Fig. 2, this time from a slightly different point of view. Are we really interested in the absolute performance of a communication system at a specific transmit power or signal-to-noise ratio, or are we actually interested in the improvement that a new system delivers over its predecessor or any other reference we have in mind? The difference between these two viewpoints changes the way we should measure and present our results.

First, the absolute values on the abscissa loose importance. We know beforehand in which region the old, known system (for example, the 1x2-system in Fig. 2) was operated. Therefore, its successor will be operated in the same region of the abscissa as the total transmit power will not change when upgrading from one system to the next one. Whether a certain throughput is achieved at a certain signal-to-noise ratio is one thing, but what is of more interest is the performance increase of the new system (or more generally, the performance relative to a known system).

Second, we always draw at least two curves: one for a system that is known and a second one for the system under investigation. At best, we choose a well-known curve as a reference so our results can be compared to others by placing the two results on top of each other in image editing software and matching them (which turned out to be a convenient way for validating our own results). Sometimes we prefer a different reference curve, namely the unconstrained mean channel capacity assuming Gaussian input signals and full/no channel knowledge at the transmitter. As this curve is only a property of the channel (and not the communication system) it is handy as a reference that is not linked to the transmission scheme being investigated. We have also plotted this curve in Fig. 2. It is the topmost dotted curve that represents what the 2x2-system could theoretically achieve. In fact, it is “quite far away” from the achieved throughput, leaving much room for improvement in the years to come.

On the other hand, focusing on relative results also influences the measurement method as we can easily reduce the uncertainty in our results by a technique often referred to as blocking [3]. The basic idea is simple: when we compare the throughput performance of a 2x2-system to the performance of a 1x2-system, we measure these two systems right after the other. Instead of first measuring the 2x2-system for several receive antenna positions and then the 1x2-system for several receive antenna positions, we make sure that the transmissions are performed over the same channels. Therefore, when subtracting the two performance curves from each other (the dashed curve in Fig. 2) the standard deviations of the error do not add. In fact, as the two throughput curves are correlated, the uncertainty of the throughput difference is less than the uncertainty of the 2x2-system throughput or the 1x2-system throughput alone (as long as the systems are not saturated). This behavior can be easily observed in Fig. 2 by inspecting the decreased confidence intervals of the difference of the two schemes.

Making all measurements relative to an initial power simplifies the measurement procedure in another way as we do not need to know the absolute level of the transmit power anymore. We only have to make sure that we can precisely attenuate the transmit signal. If required at all, the absolute scale of the abscissa can be inferred afterwards from the reference curve or a reference measurement.

The Real World Is Not Symmetric

Modern wireless communication systems employ multiple antennas at the transmitter site and at the receiver site. Unlike the situation that is all-too-frequently found in simulations, in real life, the polarization as well as the beam patterns of these antennas will never all be equal. Therefore, the expected channel attenuation of a MIMO system will be different for every transmit-receive antenna pair. More technically speaking, the channel matrix will not be independently and identically distributed.

Let us now take again a closer look at Fig. 2. In this figure, we compare the throughput of a system with two transmit antennas (the 2x2 curve) to the throughput of a system with only one transmit antenna (the 1x2 curve). In reality, it is now very unlikely that each of the two transmit antennas of the 2x2 system will deliver the same fraction of the total performance given by the whole system. For example, one of the transmit antennas might have the same polarization as the receive antennas or point more towards their direction. Consequently, a direct comparison of the 2x2 system with a 1x2 system using just one of the two possible transmit antennas would be unfair.

Fig. 4 illustrates this issue. Next to the throughput of the 2x2 system, we have also plotted the performance of a 1x2 system using the first of the two transmit antennas (the top dashed curve in Fig. 4), as well as the performance of a 1x2 system using the second transmit antenna (the bottom dashed curve in Fig. 4). Carefully note that the performance of the two different 1x2 systems is significantly different, although both are measured over exactly the same channels as the 2x2 system.

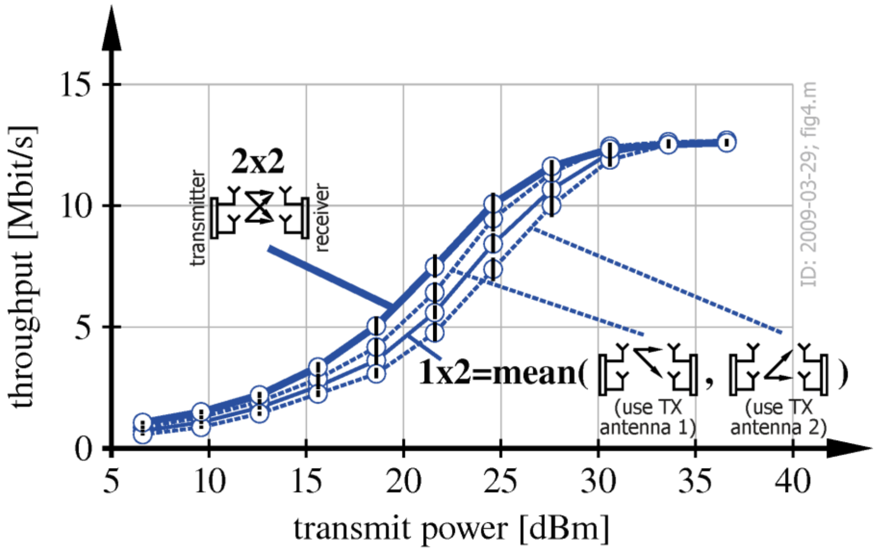

Fig. 4. When comparing a transmission with two transmit antennas (2x2) to a transmission using only one transmit antenna (1x2), we average over the two possible 1x2 transmissions to make the comparison fair.

In our experiments, to make a fair comparison between the 2x2 and the 1x2 systems, we always compare the 2x2 system to the average performance of the two possible 1x2 subsystems, both measured over the same channels as the 2x2 system (see the bottom solid curve in Fig. 4 that is also plotted in Figs. 2 and 5). By doing so, we ensure that all four possible transmission links of the 2x2 transmission are also measured equally often during the two 1x2 transmissions in which we measure two links each.

When comparing, for example, a 4x4 transmission to a 1x1 transmission, we perform one 4x4 transmission and sixteen 1x1 transmissions. In fact, we only measure four transmissions from one transmit antenna to four receive antennas and evaluate from this data the sixteen 1x4 transmissions. In simulations, we proceed similarly. We first create a 4x4 channel matrix that is not independently and identically distributed. Next, we use this channel matrix to simulate the 4x4 transmissions as well as the sixteen 1x1 transmissions. The same holds true for any other comparison. In the end, each element of the channel matrix should be used equally often to make the comparison fair.

How Precise Is Accurate Enough?

Remember, every simulation and measurement result consists of a value (the best estimate of the particular quantity of interest) and an associated measurement uncertainty (the probability distribution characterizing a reasonable dispersion of the value, visualized for example by the confidence intervals in Fig. 2). Standardized since 1993 in the "Guide to the Expression of Uncertainty in Measurement" [7], dealing with uncertainty in one way or another has always been state of the art, whereas engineers are typically only interested in the values alone.
As explained above, the results shown in Fig. 2 were obtained from 303 statistical samples *(namely independent throughput measurements)* for each transmit-power level and for each transmission scheme. The dots represent the best estimate for the mean throughput at each transmit-power level *(namely, the average throughput, as no knowledge other than the samples was available)*. The black vertical lines represent the corresponding 99.5% confidence intervals for the mean. That is, if the measurement were to be repeated, in 99.5% of the repetitions, the unknown true, average, mean-scenario-throughput would lie within these confidence intervals *(namely, the 99.5% BC_a confidence intervals obtained by bootstrapping the throughput values obtained).*

Next, remember, the precision of a sample mean in a Monte Carlo simulation can be increased by a factor of n by simulating n^2 samples (given that the samples drawn from the same distribution are independent). The same holds true for a well-designed measurement. Unfortunately, techniques such as multi-core-measuring or cluster-measuring usually do not exist. Then, the only solution to keep the measurement time at a reasonable level is a more careful experiment design and the implementation of, for example, so-called variance reduction techniques [2], [3], [4].

Still, it is not beneficial to simulate the throughput of a wireless communication system with a precision of 0.001%, even if this were possible for a given model at the expense of needing to use a fast computer, the relevance of the last digits of this result will be questionable. In other words, the conclusions to be drawn would typically not change if the result were simulated only with a precision of 1%. The simulation time, on the other hand, will change by a factor of a million.

As the old saying goes, "5% is pretty good; 1% is wrong anyway." Of course, these numbers should not be taken for granted. Still, there is more truth in them than we might think at first glance. Anyone who has ever tried to reproduce an outdoor wireless measurement in another city will see the accuracy of the results obtained with a different perspective. Even repeating a measurement indoors can be a difficult endeavor.

To test the ideas presented above, we repeated the measurement shown in Fig. 2, but in contrast, only measured a tenth of the statistical samples for each transmit-power level, that is, 30 samples (see Fig. 5). As a consequence, on the one hand, the measurement time was reduced by a factor of ten. On the other hand, the confidence intervals increased only by a factor of roughly three. Note that the conclusion that the transmission with two antennas is significantly better (99.5% significance level) than the transmission using only one transmit antenna can still be drawn up to a transmit-power level of approximately 30 dBm (as the confidence intervals for the difference do not touch the abscissa). Note again that the confidence intervals of the difference curve are much smaller than those of the other curves as the latter are correlated. Of course, 30 samples are only sufficient if we are interested in these relative results. But frequently, only relative results are of interest.

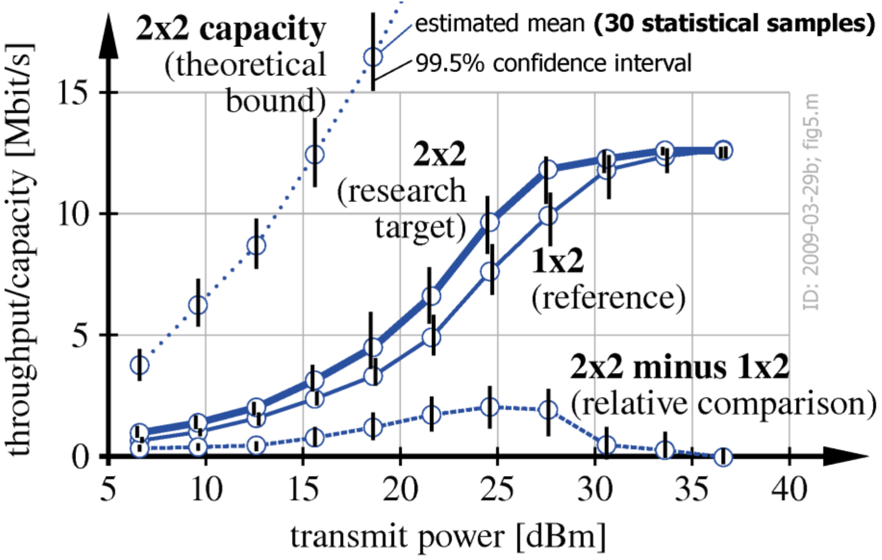

Fig. 5. The same scenario as in Fig. 2, but only a tenth of the statistical samples were measured this time. Note that the confidence intervals for the relative comparison are smaller than those of the 2x2 curve and the 1x2 curve that already overlap. Compared to Fig. 2, the confidence intervals are about three times larger, the measurement time is reduced by a factor of 10, and the conclusion (the relative comparison is greater than zero) does not change.

Buying Expensive Black Boxes

Once the initial simulation is finished, the next step consists in browsing the catalogs of different vendors and starting to spend lots of money on “fancy” hardware. But wait! There is one thing that we have learned during the last years the hard way: hardware does not work out of the box!

So we should rethink our strategy and carefully consider *how* to buy hardware for a wireless communications test bed and what important points need to be considered.

Let us, for example, take a closer look at the digital baseband hardware. Such hardware can be bought from a variety of vendors in very different price ranges. In contrast to standard personal computer parts, for example, such specialized hardware is sold in very small quantities. Sometimes, it is even assembled directly for the paying customer. Still, that does not hinder most companies in this field in pushing out new and more powerful products on a regular basis. So how is it possible for these hardware manufacturers to keep up with ever-shorter product life cycles?

On the one hand, the price for digital signal processing hardware in wireless communications is premium, as it has to include not only a large portion of the company's profit but also of the development costs. On the other hand, *you* will be the alpha tester, typically even for demo examples and user manuals. The effort required to make such hardware work properly is often substantially underestimated. In our experience, if good support is not included and you do not have a lot of time to debug somebody else's hardware, there is little you can do.

Recently, the advent of reprogrammable hardware enriched the quest for buying hardware that really works by so-called firmware upgrades. A really wonderful idea in principle, but the possibility of firmware upgrades did not actually increase the quality of hardware but rather actually decreased it. The reason is very simple. Before the concept of firmware upgrades existed, a company had to take back a product if it did not work properly, which implied high costs. Now, all a company has to do is to release a firmware upgrade on their homepage. Continuing this train of thought, companies *could* now even sell reprogrammable products that are not thoroughly tested or even refuse to work at all. Looking at the bug-fix lists of certain products, no other conclusion seems reasonable for us besides that a product which such long bug-fix list could have never worked flawlessly at the time it was introduced to the market. Fortunately, this argument can also be turned around. In our experience, hardware that does not allow for firmware upgrades tends to be more robust than hardware that allows for firmware upgrades. While experience has shown that this hypothesis might be true, the reason for this could also be that products without firmware are just simpler and therefore do not fail as often. Still, the general trend is undeniable. When buying a product that allows for firmware upgrades, one should be prepared to wait for the updates and even to fight for them (and based on our experience, this can take years, and holds true for almost every product).

Summarizing, one has to be prepared to buy hardware with inputs and outputs and little to no documentation on the inside. In other words, one has to be prepared to buy expensive black boxes that allow for firmware upgrades and typically do not work out of the box (see Fig. 6).

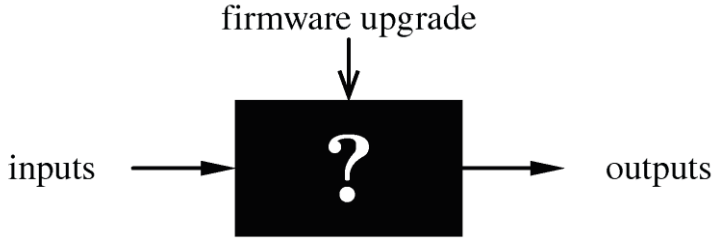

Fig. 6. Every piece of hardware bought is more or less a black box. It has inputs, outputs, and the possibility of a firmware upgrade, but little or nothing is documented about what is inside.

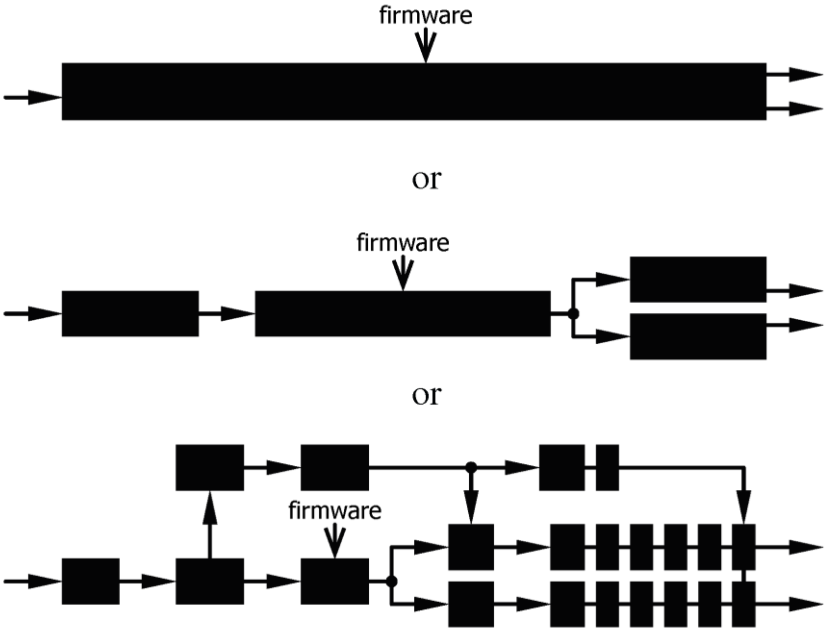

Fig. 7. “The black box problem”: should we buy the hardware in one complex piece, in a few pieces of moderate complexity, or in a lot of very small but simple pieces that can be easily checked and repaired?

Knowing about this, the key task in buying hardware for a test bed is not only to choose the right black box from the right vendor, but more importantly, the right black box size (see Fig. 7):

- One way is to buy the hardware in one complex piece: At first glance, this also seems to be the least-time demanding task. All that has to be done is to search for an appropriate piece of usually very powerful and flexible hardware that suits the desired purpose, next to maybe writing an application for funding this piece of hardware. However, this approach will be prone to fail for the reasons outlined above: Almost certainly, the hardware will simply not work as desired out of the box. Unfortunately, a researcher is then at the mercy of the hardware manufacturer to release new firmware. Trying to correct the problem will, on the other hand, typically also fail because of poor documentation.
- The other extreme is to assemble the hardware from many small and simple pieces. Consequently, these small pieces of hardware are relatively cheap. They can be bought

> from a variety of different vendors. And most important, it is possible to exactly specify their functionality and interfaces prior to purchase. The implications for the time after purchase are straightforward. First, it is relatively easy to understand and check the functionality of the hardware bought. Second, it is also relatively easy to prove to the vendor by measurements that the hardware does not fulfill its advertised specifications. Therefore, experience has shown that vendors will (have to) correct their products. Third, if a product does not live up to its expectations, buying a better-suited product from a different vendor is a suitable option. Fourth, if a product does not work as advertised and the vendor refuses to repair it for obscure reasons, writing it off and buying an alternative product may be preferred over any legal actions. Fifth, because it is less complex and, therefore, often more thoroughly tested, small pieces of hardware are less likely to fail in general. Finally, the different pieces of hardware can be reused for different experiments, assembled in a different way, and so on. In other words, the investment made in a small piece of hardware is not lost after the measurements which it was purchased for are finished.

Note that we do not consider a carrier board with several different modules interconnected by proprietary interfaces as modular. In our opinion, this is one black box featuring interchangeable modules that might fail because it is very unlikely that they have been tested in the way you will assemble them. Conversely, a rubidium frequency normal connected to an oscillator via a 50 ohm coaxial cable (a well-defined standard) or a power meter connected to a PC via a local area network connection (also, a well-defined standard) is very likely to work as expected.

On the downside, assembling small black boxes to obtain a complex piece of hardware takes considerable time, expert knowledge, and expensive (measurement) equipment for testing. There is absolutely no need to redesign things that others have designed, such as a GPS receiver where only its outputs are of interest and it is very unlikely that “half a GPS receiver” might be needed in a future experiment.

The art of efficiently and effectively buying hardware for a measurement is now, in our opinion, to maximize the size of the black boxes purchased while still being able to fully specify and test their functionality and to maintain flexibility for both, present and future experiments.

Real-Time, Off-Line, or Even Both?

Programming digital signal processing algorithms into real-time hardware such as digital signal processors and field programmable gate arrays requires time. Debugging and testing algorithms programmed into such hardware requires even more time. Finally, modifying these algorithms once they are inside the hardware also seems to be very time consuming. Even worse, people researching algorithms are usually not trained and experienced in programming real-time hardware. Even if they were, the tools available are, at least in our opinion, not the most convenient and stable. As long as researching on how to implement algorithms in real-time hardware is not needed or of interest, there is no real benefit in doing so other than decreased execution time (that can be combated with cluster processing on standard PCs) and publicity (that can be restricted to a few, simple, nice-to-show setups).

On the other hand, coding algorithms in a high-level programming language is convenient. This is especially true when utilizing a numerical computing environment such as MATLAB that comes with a lot of toolboxes. Furthermore, as such an environment may have been used to test the algorithms in an initial simulation, the idea of reusing this code also for a measurement seems obvious.

Consequently, in our measurements, we avoid programming algorithms into real-time hardware at all cost. Instead, we execute all algorithms off-line on personal computers and use the same high-level language code that has already been used in the initial simulations.

To do so, we prepare a transmission in which:

- We organize all signals to be transmitted in blocks.
- We always transmit the same data and ensure by simulation that the measurement result, namely in our case the throughput, is not altered by this simplification of reality. Furthermore, we design the experiment in such a way that only a limited number of different blocks is need to be transmitted during the course of the whole experiment. Whether or how often these different blocks will be transmitted in the experiment can be decided later; for now, we only have to decide on a finite set of blocks.
- We calculate the baseband data samples of all these different blocks and store them on fast hard disk drives or in the memory of the transmitter

Next, we carry out several transmissions during an experiment in which:

- We load the block to be transmitted, namely, the baseband data samples, into a real-time capable buffer.
- We transmit this data block in real-time over the wireless channel.
- We receive the data block into a real-time capable buffer.
- We store the data block on a hard disk drive (rather than evaluating it immediately).

Finally, we evaluate the stored data blocks off-line in a cluster of personal computers.

Having the data readily available at a transmitter that is synchronized with the receiver not only allows for the transmission of single blocks of data. Interestingly, it also allows for techniques such as retransmissions (see Fig. 8), adaptive modulation and coding (see Fig. 9), and adaptive modulation and coding combined with pre-coding and retransmissions (see Fig. 10). The only requirement to do so is a channel that is static during one measurement or a non-static channel that can be repeated.

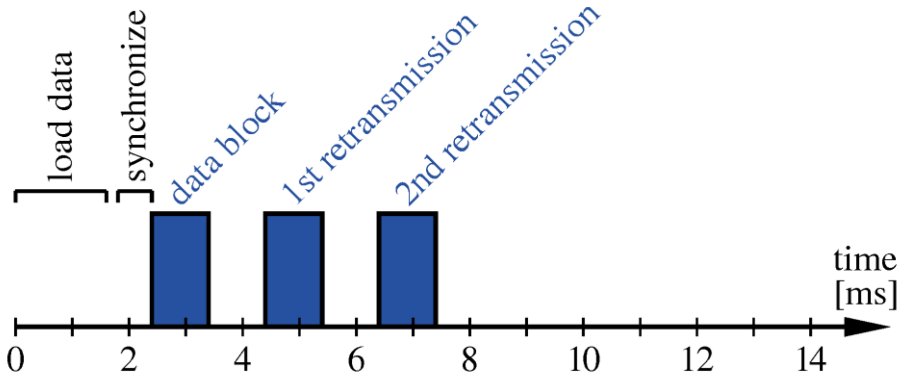

Fig. 8. We always transmit all possible retransmissions even if they are not required. Later, during the off-line evaluation, we decide whether we need them or not.

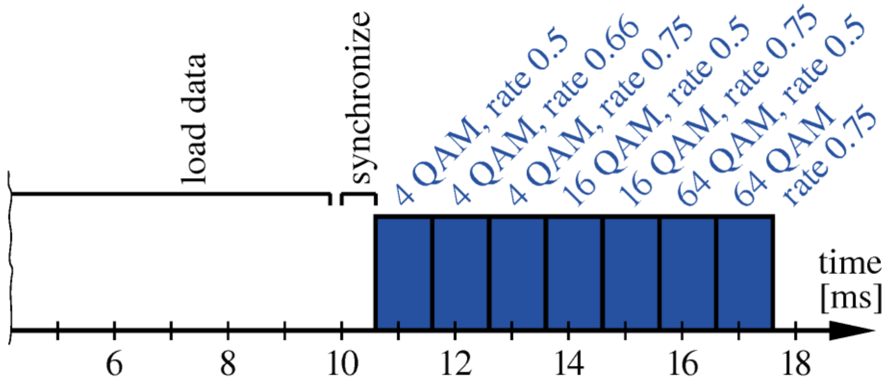

Fig. 9. We always transmit all possible adaptive modulation and coding schemes (faster than the channel coherence time) even though only one would be transmitted in a real-world transmission. Later, during the off-line evaluation, we decide which one to choose based on the estimated channel.

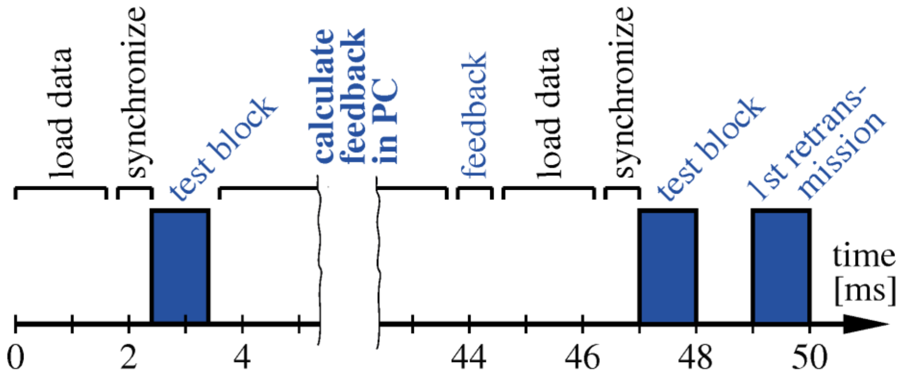

Fig. 10. 0. If the number of possible schemes exceeds the channel coherence time, we first transmit a training block. Next, we let the receiver evaluate the channel conditions (but not more) and estimate which blocks should be transmitted under these channel conditions. Then we feed this information back to the transmitter and transmit the blocks selected (with the retransmissions that may be required or not) by loading pre-generated blocks from the memory or fast hard-disk drives. Finally, we evaluate not only the whole measurement offline, but also if the channel has changed between the training block and the actual transmission. If the channel has changed, we discard the result of this transmission.

The Wrong Metric?

Compared to other metrology related issues, building complex measurement setups in wireless communication requires considerable experience and engineering knowledge. Because the art of engineering setups has not been widely covered in scientific literature, one may only be able to find books on certain areas of interest, for example, for designing radio frequency hardware [8], [9]. Besides such books, the literature on how other groups have set up their hardware for measurements is either limited or non-existent. So carefully note that this will also apply to your "research."

Let us take a look at, for example, the steps required for synchronizing several transmitters and receivers in a wireless communication system measurement. This issue becomes non-trivial if measuring outdoors, over large distances, or at high velocities. Sooner or later, every research group will have to spend a considerable amount of time and manpower in developing and testing an engineering-solution that serves no other purpose than synchronizing the measurement set-up. Then, sometime later, they may find that there are new features required by this hardware and that the whole set-up must be redesigned. Comparing the manpower and time invested to the scientific output achieved by this particular synchronization hardware (that is, zero), it would have been more efficient to buy somebody else's design and to copy it. Unfortunately, typically this possibility does not exist. In addition, governmental funding makes it often more complicated to spend one man-year on developing mid-priced hardware from scratch rather than buying it [10].

Summarizing: In order to make measurements, a lot of time is “lost” that could have been otherwise devoted to research. On the other hand, it is not really lost, because it allows us to measure.

Some Years Later...

Did we not mention it at the beginning? Measurements cost time and money. In return, they provide us with invaluable insights into real-world problems. Whether a paper could have been published more easily by submitting only the initial simulation and skipping all the measurement effort is another (very sad) story.

References

[1] S. Caban, C. Mehlführer, M. Rupp, and M. Wrulich, *Evaluation of HSDPA and LTE: From Testbed Measurements to System Level Performance*, 1st ed. Wiley, 2011. ISBN: 0470711922.
[2] R. Fisher, *The Design of Experiments*, 1st ed. New York: Wiley, 1935.
[3] R. L. Mason, R. F. Gunst, and J. L. Hess, *Statistical Design and Analysis of Experiments*, 2nd ed. Wiley-Interscience, 2003, p. 221. ISBN: 0471372161.
[4] W. G. Cochran, *Sampling Techniques*, 3rd ed. Wiley, 1977. ISBN: 047116240X.
[5] B. Efron and D. V. Hinkley, *An Introduction to The Bootstrap (CRC Monographs on Statistics & Applied Probability 57)*, 1st ed. Chapman & Hall, 1994. ISBN: 0412042312.
[6] A. M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*, 1st ed. Cambridge University Press, 2004. ISBN: 052183127X.
[7] *Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization (ISO)*, 1993.
[8] A. Behzad, *Wireless LAN Radios*, 1st ed. Wiley & Sons, 2007. ISBN: 0471709646.
[9] A. Luzzatto and G. Shirazi, *Wireless Transceiver Design*, 1st ed. Wiley & Sons, 2007. ISBN: 047006076X.
[10] S. Caban, A. Disslbacher-Fink, J. A. García-Naya, and M. Rupp, "Synchronization of wireless radio testbed measurements", in *Proc. of International Instr. and Meas. Technology Conf.* (I2MTC 2011), Binjiang, Hangzhou, China, May 2011.

Bios

Sebastian Caban is a postdoc researcher at the Vienna University of Technology, Austria. In this productive environment he is now already living the art of building testbeds for wireless radio communication systems for more than six years seven days a week. Before, Sebastian Caban studied electrical engineering at the Vienna University of Technology and the University of Illinois at Urbana Champaign to receive his MSc and PhD degree as well as business administration at the University of Vienna to receive his BBA and MBA degree. He is a member of the IEEE and IMS.

José Antonio García Naya is an assistant professor at the University of A Coruña, Spain. He once came to Vienna to learn the art of wireless engineering and finally came back several times to continue with the fruitful research collaboration that started at his first visit. Before, José Antonio García Naya studied computer engineering at the University of A Coruña to receive his MSc and PhD degree. He is a member of the IEEE and IMS.

Markus Rupp is a full professor at the Vienna University of Technology, Austria. He is spending most of his time with bureaucracy and raising funds, that is, enabling researchers to build wireless testbeds for experimental research. Before, he studied electrical engineering at the University of Saarbrücken to receive this MSc degree and at the Technische Universität Darmstadt to receive his PhD degree to continue then at the University of Santa Barbara and at Bell-Labs. He is a senior member of the IEEE and past president of EURASIP.