40GBase-T

by Dr. Franz-Joachim Kauffels

For over 20 years, Dr. Franz-Joachim Kauffels has been one of the most experienced and best known reporters of the network scene. He has written over 20 textbooks and innumerable articles and is well known for his lively and stirring seminars. In addition to his work with ComConsult Akademy, he is an independent strategic consultant with focus on LAN, operating software, connectivity and network management systems. He is member of several scientific work groups and worked as lecturer for economic informatics at the University of Essen.

Experience shows that a new step forward in speed will only be widely accepted by the market if there is also a twisted pair version. Until recently no-one could say whether a 40 Gigabit version of Ethernet was also possible over twisted pair. That has now changed and there will definitely be 40 GBASE-T. Even if an appropriate standard has to wait for a few years, we can already say with certainty what kind of cabling is suitable for this purpose. Given the life of cabling systems, that is a very important answer. In this article we will obviously also be looking at the possibilities for transceiver technology.

There have been standards for 10 GbE over fibre since 2003. At first few users showed an interest in them. This only changed with the availability of a twisted pair version, 10 GBASE-T. Although many operators are finally leaning toward fibre with the introduction of 10 GbE, the mere existence of a twisted pair version seems to offer considerable reassurance. The CX versions over Twinax cable are admittedly technically excellent and are often used, simply because manufacturers like HP incorporate them in the Switch Blades of the Blade Server, but these are far from having the same psychological effect.

The burning issue now is obviously whether there will also be a 40 GBASE-T. Until late autumn 2008 nobody could answer this question. Then scientists at Penn State University succeeded in making the breakthrough: 50 Gigabits per second over 100m of Category 7A twisted pair cable!!!

This meant two things of tremendous importance for planning:
- there is a twisted pair cable that is suitable for 40 GBASE-T
- there is a suitable connector

This result plays a vital role in planning. Even if there is currently no 40 GBASE-T standard, and even if 40 GbE will not be used in the foreseeable future, future-proof planning of the cabling is already possible. This is tremendously important, given the long life of cabling systems.

However, this result is also particularly important for standardisation. In particular, when drafting the standard for the transceiver and its properties, it is possible to refer to the already established characteristics of the cable or of the physical transmission channel that it forms. The standardisation of 10 GBASE-T was so prolonged, because there had not until then been any connection of that type. Thus there was a multiplicity of alternatives for different real or imagined cable specifications within the standardisation process. This had the consequence that the standard for 10 GBASE-T was established first, only then followed by the Cat. 6A cable to go with it.

However, it must be noted at once that there is a considerable difference between the standardisation of 10 GBASE-T and that of 40 GBASE-T. With 10 GbE a lot of time was wasted developing a standard for unscreened UTP cable. With 40 GbE it is obvious from the outset that there can only be a solution for screened STP cable.

The Cat. 7A cable is also currently in the process of standardisation. However, some manufacturers are already offering such a cable. In what follows we refer to Nexans' LANmark 7A product.

The properties of the 7A cable are defined up to 1000 and/or. 1200 MHz – so far beyond the range defined previously. The standards only define Cat 6 cable up to 250 MHz, but Cat 6A standard compliant products are specified up to 500 MHz.

The basic data on the 7A cable is:

- NEXT (Near end cross-talk) cancellation: 60 dB at 1000 MHz

- FEXT (Far end cross-talk) cancellation: 50 dB at 1000 MHz

- RL (Return Loss): 8 dB at 1000 MHz

- ANEXT (Alien cross-talk) cancellation: 0 dB at 1000 MHz (!!!)

ANEXT is also often known as "Alien Crosstalk", The LANmark 7A cable is fully screened (obviously only if properly installed), so that the "Alien" noise cannot cause any further damage there.

These are values that put the cables we have been familiar with up to now very much in the shade. For a comparison see Fig. 1.

The value of ANEXT is sensational. One difficulty with 10 GBASE-T was to preserve what was for a cable a relatively delicate signal from external disruptive influences. In the initial versions it was possible for the signal to be lost completely. This quality was achieved by means of an S/FTP design with four shielded pairs and an overall screen.

Now we need a suitable connector. The GG45 connector lends itself to this. It has 12 pins. The "2-in-1" connector combines RJ45 and GG 45. This results in two modes:

- RJ45 mode up to 500 MHz for 1 and 10 GBASE-T

- GG45 mode up to 1000 MHz for 40 GBASE-T

The GG45 connector is suitable for Cat 7A cables. It is already fully standardised as ISO/IEC 60603-7-7. The RJ45 mode is mandatory under ISO 11801 to guarantee backward compatibility (see Fig. 3).

An alternative connector would be the IEC 61076-3-104, developed by Siemon. This has a new pin layout and is therefore not backward compatible. Apart from that, the result mentioned at the beginning was obtained with a combination of the LANmark 7A cable and GG45 connector. In what follows we refer exclusively to this.

	LANmark-7A	Category 6A
→ Cat. 6A specified for 10 GBASE-T → Cat. 7A under discussion, but there are already points, for example Nexans LANmark-7A		
• NEXT	60dB at 1000MHz	30dB at 500MHz
• FEXT	50dB at 1000MHz	25dB at 500MHz
• RL	8dB at 1000MHz	8dB at 500MHz
→NEXT: Near end cross talk cancellation →FEXT: Far end cross talk cancellation → RL: Return Loss		

Fig 1: Cat. 6A and Cat. 7A

$\rightarrow$ S/FTP design with four individual screened pairs
$\rightarrow$ Performance up to 1200 MHz
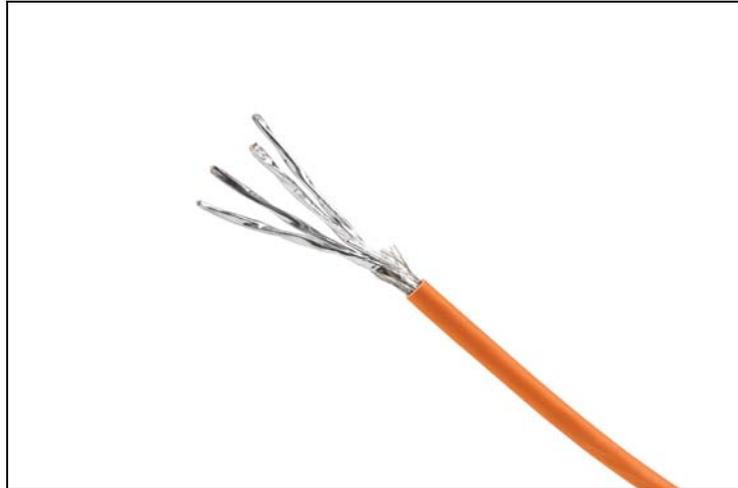$\rightarrow$ NO(!!!) Alien cross-talk

Fig 2: Cat. 7A Cable

Applications for 40 GBASE-T

What possible applications are there for 40 GBASE-T? For the time being, expansion of the market for 40 Gigabit Ethernet is going badly. It was ultimately also the case with Gigabit Ethernet and 10 GbE that the market only really picked up when the copper versions were available. Funnily enough, large quantities of fibre interfaces were then bought. The market is not really logical, but its fundamental mechanisms are quite easy to understand. I have already reported on the possible applications of 40 Gigabit Ethernet on a number of occasions, but obviously there is a painful gap in the standard despite the many PHY versions. If you look at it closely, there are interfaces for bridging hundreds of kilometres, but only comparatively few that you can actually use in the data centre. And when we are talking about 40 Gigabit Ethernet, we are also talking about 40 Gigabit Fiber Channel, since this standard does use a different packet format, but otherwise is currently aligned in its physical interfaces with the 10 GbE definition, albeit partially with a small form factor. There is no reason why the FC working party should deviate from this hitherto very successful strategy. Finally, we will also discuss FCoE, as it is technically identical in any case.

So where in the 40/100 Gigabit standard is there now a solution for interconnecting servers with one another or linking servers to storage devices really cost-effectively using existing cabling? The four-channel 40 GBASE-LX-4 version for the use of 10 km single mode fibres is not really satisfactory for the data centre and the generally "forgotten" Twinax 40 GBASE-CX-4 version with its 10 metre range is not enough. The 40 GBASE-SR multimode version has no electronic dispersion compensation as standard. However, manufacturers like AMCC are already able to achieve this – so sooner or later there will also be a 40 GBASE-LRM version by analogy with 10 GBASE-LRM. That would be the only suitable data centre version. If you are investing millions in SAN, the purchase of a few metres of new fibre no longer

matters. But what happens in the many other cases in which e.g. you want to connect a single server 85 m away up properly? Where you want to connect SAP application servers to one another? Where you want to connect up one of those super new tape decks? What happens when you "suddenly" notice that virtualisation means that your 10 GbE solution is no longer sufficient? It is indeed nice of the component manufacturers to be giving us the prospect of cost-effective transceivers and server boards right now, but that is only really of assistance to possessors of structured fibre cabling with the "right" fibres. Thus, give or take an application area, without an obvious copper version the standard for 40 Gigabit Ethernet appears rather incomplete and currently uses mainly metropolitan network providers. Since I am still somewhat cross about this, I should like to add that the standard in its present form hinders as much as it helps the creation of a high performance network infrastructure since, because of the lack of a useable copper version on the market (and this cannot be replaced by a low-cost fibre version), there is still a perceptible coolness that in the last analysis is leading to distinctly unsound and expensive scrabbling about for other solutions. Looking a little ahead, as with 10 Gigabit Ethernet, the overwhelming majority of the installed base of 40 Gigabit Ethernet is to be found in data centres. Here people would like to support the installed base of computers simply by using the installed base of structured cabling better to give higher performance. This embraces not only server to server or server to storage device links, but also the route to switches and DWDM systems for implementing remote connections. For at least two years, people have been talking about growth in 10 Gigabit interfaces for servers.

It is simply the standardising power of reality that is leading to an autosensing 1000/10000 chipset generation and only out of consideration of volumes that chip manufacturers concern themselves, independently of this, about whether a user actually needs such a data rate. It is simply a matter of price and volume. Faced with this development, the only thing left is to apply the most brutal of overkill to horizontal and building backbone cables. As with shared medium systems, people blithely hope that all users enjoy sleeping and will not latch on to the idea of using the available performance up to the terminal. On the contrary, people are installing Cat. 7 cable so that the Gigabit also gets to the next router in good order, connects horizontal cables to one another, but also only with a Gigabit link because there is nothing else that they can afford.

GG45
Revolutionary 2-in-1 connector technology
Fig 3: GG45 connector Photo: GG45 Alliance

Another standard that – despite all the doom mongering – is in the process of slowly but surely spreading is iSCSI, the mapping of storage block transfers to IP packets and networks. If we look at the performance limits of current devices and extrapolate them in accordance with Moore's Law, as is so common, we will have a ten-fold increase in performance within 4-5 years. Suddenly we have something on the net that wasn't there before: fileserver to fileserver, fileserver to application server and fileserver to backup media traffic. This will send quite a few gigabits up in smoke! Then we come across a couple of wireless VLANs, where for each cell, so for each 11n access point, you will soon also have to reckon on 500 Mbit/s and for high-performance cells this is none too big – so we get a grand total of hundreds of cells, each with 4-8 users, and once again we have conjured up a backbone performance of 10 gigabit. Forgive me the flippant style, but the figure doesn't get any less when put more formally.

These are things that you know. But they aren't everything. In the foreseeable future, particularly in the course of virtualisation, I am reckoning on the following tendencies: hardware supported CPU load balancing will no longer be limited to processors that happen to live in one box but also extend at least to neighbouring computers over high-speed networks. For this we need reaction times in the sub-microsecond range and any amount of raw performance. There will be "TCP/IP Offloaders" that do nothing but take over all the tasks that crop up in the TCP/IP environment. Since we do everything with TCP/IP, that is a lot of work. The net processors can execute TCP/IP completely on one chip, and hence it is no longer acceptable if e.g. a server has to think too long about TCP/IP. We are getting a foretaste of this with iSCSI hardware accelerators. Overall, the processing speed of protocols by hardware processors and offloaders will increase dramatically, so that a network can no longer rely on the communicating searches recalculating the protocols so ponderously that they can only spew out data relatively slowly and with long pauses. Finally, there are such nice novelties as RDMA, the Remote Direct Memory Access protocol, which, as its name suggests, supports DMA on another computer.

Goals for standardisation of 40 GBASE-T will be:

- Bridging a distance of 100 m on Cat 7/7A or higher twisted pairs

- Conservation of existing investment of cabling in data centres and structured cabling elements

- Support for 10 GBASE-T and 40 GBASE-T with a single PHY with autonegotiation as part of "scalable Ethernet"

- Support for the 40 Gigabit XLAUI interface
- Multiple PHYs for higher speeds with trunking.

These are associated with the normal goals of Ethernet standardisation, such as:

- Preservation of the 802.3/Ethernet frame format at the MAC interface

- Preservation of the functional requirements of 802 with the exception of the hamming distance

- Preservation of the minimum and maximum frame sizes

- Support for full-duplex operation alone

- Support for star network topologies with point-to-point connections in structured cabling environments

- Specification of an optional Media Independent Interface (MII)

- Support for P802.3ad link aggregation

- 40,000 Mbps at the MAC/PLS service interface

40 Gigabit on Twisted Pair and Shannon's laws

In Fig 4 we see the configuration we are aiming at: from 40 Gigabit to 40 Gigabit Transceiver over 90 + 5 + 5 = 100 m of STP cable, 8 conductors in 4 pairs as in 10 GBASE-T, actually "just" an upgrade of the latter.

In order to achieve this, we have to think a little. Fundamentally, Shannon's laws apply, according to which a maximum of two signal transitions per second can be packed into an available bandwidth of 1 Hz.
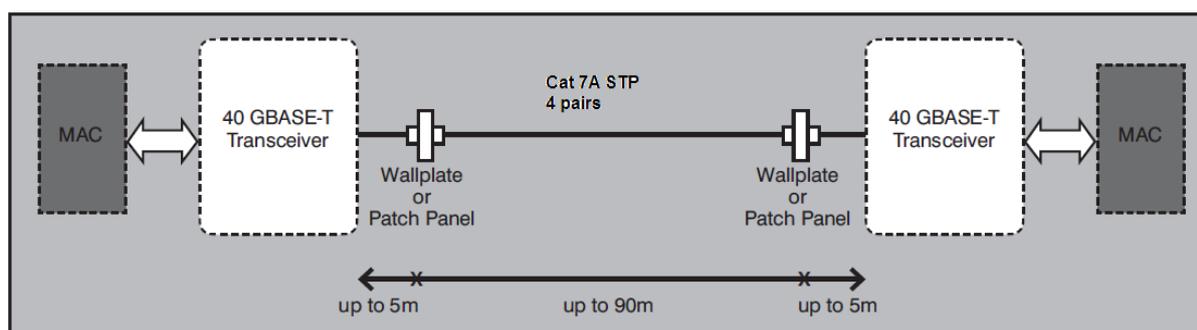
Fig 4: Configuration

Although you can get by quite easily with 10 megabit Ethernet, at 100 megabits people were already thinking about this and using a 3-value encoding, which led to a cable with a necessary bandwidth of 33 MHz. Gigabit Ethernet over Twisted Pair uses a 5-value coding in which 8 data bits and 1 control bit are converted to 5-value signals by means of Trellis coding. Linear binary line coding would need a bandwidth of 500 MHz for gigabit Ethernet, but this can already be divided into 4 communication technology independent channels, so that each channel only needs a bandwidth of 125 MHz and, because of the multi-value transmission, ultimately only 67.5 MHz per direction. With full duplex operation, this ultimately boils down to 125 MHz per pair of conductors. However, this concept also has its limits. If we wanted to implement the same strategy as with Gigabit Ethernet, there would be 320 Bits instead of 8 in the encoder. Four 80 bit groups bring this up to approx. 4 million different states per group. A substantially higher order logic does not get us any further.

Shannon's limits are not in any sense dependent on modulation techniques, the reverse is true: they represent a measure of the merit of a modulation technique, since the better a technique is, the nearer it approaches the theoretical limits.

If we look at 1000 BASE-T a little more closely, we will observe that there are a series of assumptions that were made at the time of the definition and remain there completely unchallenged to this day. They involve, for example, attenuation The bandwidth available for transmission is assumed. And this assumption is laid down by e.g. a cabling standard. Many owners of token ring networks assume that the cable only offers 16 Mbit/s until they are taught better. Furthermore, it is assumed that there are irreducible sources of noise, such as e.g. background noise, cross talk from other cable pairs, alien noise and noise from the transceiver. All these are fine as assumptions but the reality may prove to be slightly dfferent. This means you can mostly get by unobserved because most people are not in the least interested in what is actually happening but merely in an indication of how far they can run the cable so that it still works. This situation is not new and also exists with wireless systems and optical networks. The market is simply crying for a system of boxes; what they want is not thought but tick boxes.

And if you want to transmit 40 Gigabit/s on 8 twisted pair conductors, you have to rethink things substantially and check the facts.

Fig. 5 summarises the disruptive influences once again. Near-end crosstalk NEXT from adjoining cable pairs, far-end cross-talk FEXT from the transceiver and adjoining cable pairs, general cancellation and alien crosstalk, i.e. also electromagnetic Interference.

A categorised cable must be of high quality with low material-determined variation in order to meet the requirements of TIA-568. The standard gives bandwidths at which certain characteristics must be achieved. However, this is only a paper requirement, as modern cables achieve the same performances at higher bandwidths. This depends mainly on the transmission geometry and the properties of the material. Minor structural variations and irregularities in the connectors can impair these values, but with modern systems not really seriously. These days, when a supplier guarantees you safe operation of the cable at, say 600 MHz, that is a value that will be achieved under the most unfavourable conditions. For this to work, all components must be much better and must certainly achieve 1.5-2 GHz bandwidth under normal conditions.

That is why we sometimes get such labile limits at which a connection should actually no longer work according to the standard but does in practice with a system from a manufacturer.
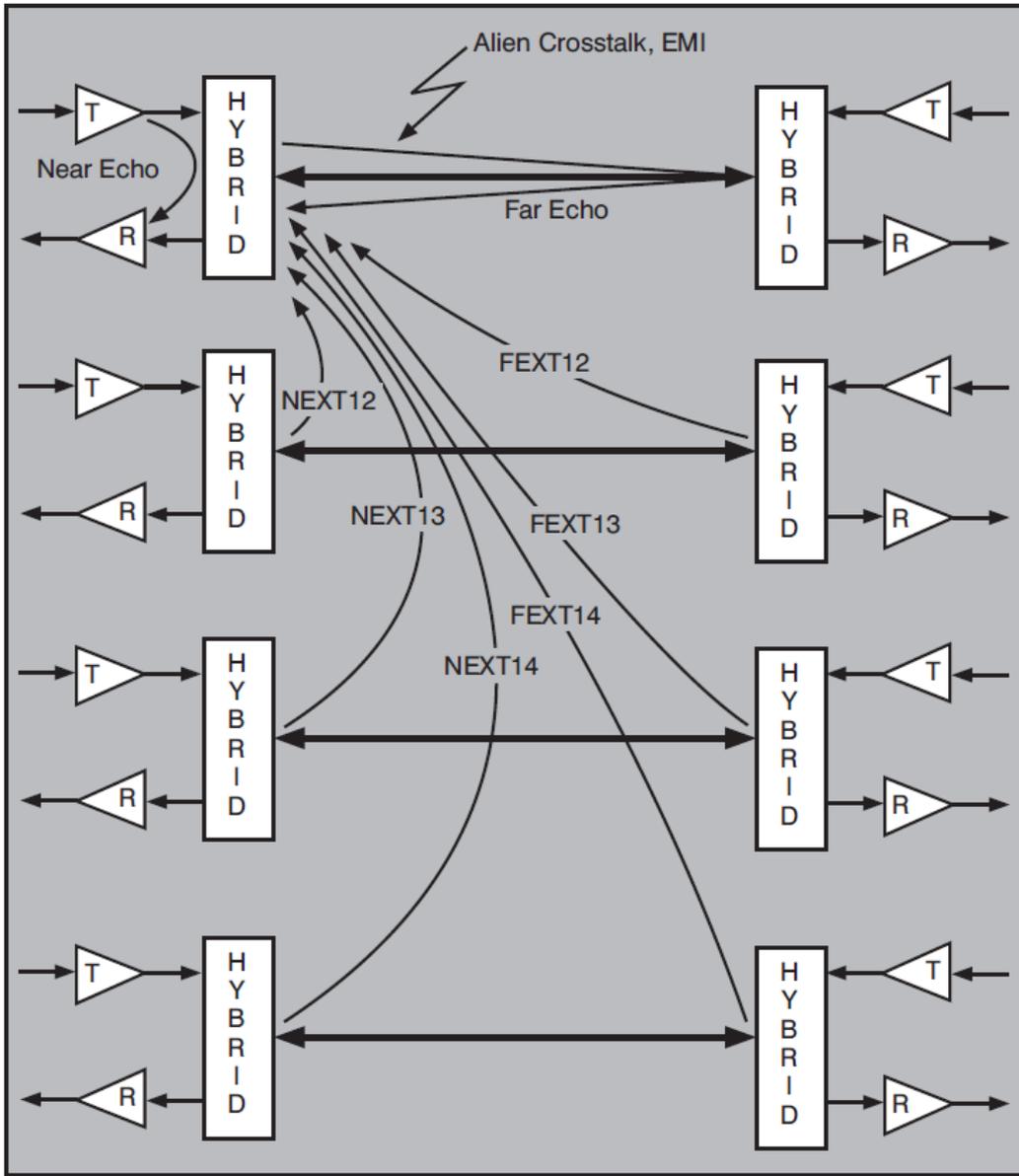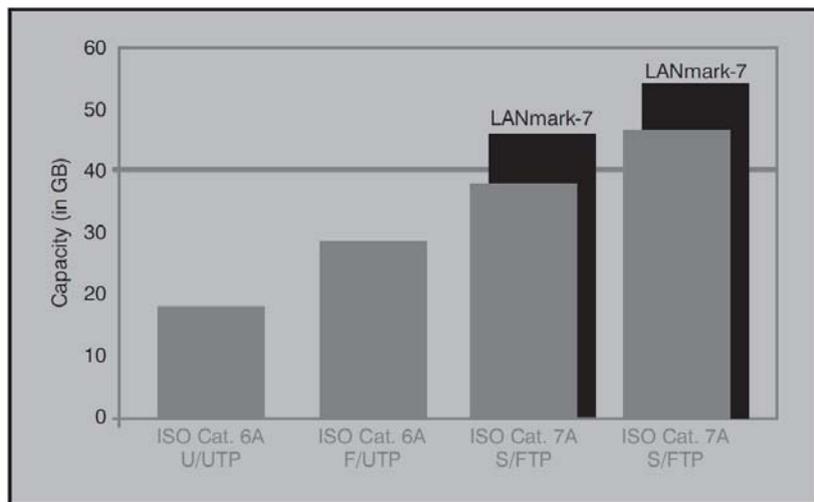
Fig 5: Disruptive influences, overall

Fig 6: "Shannon capacity" for 4-pair cable,   Source: Nexans

Thus we can work out, for example, that ISO cat 7 is simply no longer sufficient for 40 GbE operation. There are, however, Cat 7 systems from manufacturers that can nevertheless achieve this transmission rate. The next diagram demonstrates one such, again in relation to the manufacturer Nexans (see Fig 6).

Obviously the development of 40 GBASE-T can profit hugely from the results of the 10 GBASE-T standardisation, which in part came about with great difficulty. For instance, it is known that beyond 200 MHz the near-end crosstalk cancellation increases, in the worst case, by approx. 8 dB and does not rise to some stratospheric level. Nothing else should have been expected, as near-end crosstalk arises from an inductive effect that arises because the pair of conductors in question picks up an electromagnetic field that is generated by the other pair of conductors. However, at a constant distance, the power of the electromagnetic field decreases as the frequency rises. This compensates for the intrinsically increasing "sensitivity" of the other pair of conductors to this sort if interference as frequencies rise. Furthermore, this comparatively trivial 8 dB can be completely compensated out if the input power supplied to the "interfering" cable is reduced by around 1-2 dB.

Far-end crosstalk cancellation, in particular, is substantially less at e.g. 1000 MHz than at 200 or 500 MHz. As with near-end crosstalk cancellation, this has to do with the loss of intensity of the "interfering" signal at higher frequencies, as already explained. This renders extrapolation of the standard values unnecessary. Hence the 50 dB FEXT value of the LANmark 7A cable is less an achievement of the manufacturer than of the physics as such.

These effects cancel one another out to some extent. Cancellation through the total power of all cross-talk cancellation is summarised by the expression "NEXT Power Sum". Between 200 and 400 MHz there is an astonishing plateau, and only toward 500 MHz do we record any increase truly worthy of the name. However, this "collective measure" has a big effect on determining performance, as in reality the effects in question never occur in isolation, but always together.

Another important question is whether operation at higher frequencies means that the cable will cause interference with other wireless services. Over and over again I see the worried faces, fearing that the operation of 40 Gigabit Ethernet over twisted pair could interfere with the sacred wireless cells. This is definitely not the case because there is only a fully screened cable available for 40 GBASE-T. Incidentally, interference could only occur if a harmonic of the operating frequency of the cable had a frequency in a WLAN band, and furthermore was of sufficient intensity. So that readers read this article right to the end, I will not work it out at this point, but as already stated the intensity diminishes greatly

with frequency and the risk of a "threat" from a harmonic of 500, 1000 or 1200 MHz is less, or at most exactly the same as the risk of a threat from a harmonic of 66, 125 or 250 MHz. But, as I said, the screen means that there is no effect of that kind anyway.

Conversely, we can also see what happens when the screen is not implemented, breaks down or is defective: the 10 or 40 GbE useful signal would remorselessly get lost in the general radio interference! Background interference (noise) is actually practically independent of frequency.

However, there are other interesting interrelationships. As suggested earlier, in the past improvements in the use of bandwidth were achieved mainly by generating ternary (Base 3) or quinary (Base 5)signals and sending then over the cable instead of a binary line code. This makes it possible to accommodate more bits per second, usually specified in the unit Baud, in one signal transition. Everything has its price, however. With multi-value transmission the logical values are closer together and hence more susceptible to interference and it is necessary to take more trouble over the receiver, in order to keep them separate properly and decode them correctly. Finally the signal/noise ratio is once again vital. After a little calculation we come up with the result that two logical levels require a signal/noise ratio 6 dB higher. This is a relatively abstract result and many people cannot come to terms with it. However, in Fig 7 we see a rather different representation. This considers a transmission path with Cat 5/5e cable that, on the parameters given above, is endowed with better cancellation of the determinable parameters such as NEXT and FEXT. The graph shows how many bits can be accommodated per second per signal transition in the specified conditions at a set maximum transmission frequency. We observe with horror that this gets less the higher the frequency. This severely dents our hopes of getting anything out of a further substantial increase in the logical signal transitions.

In itself this result is not unexpected, since overall interference increases with frequency and we have improved the signal/noise ratio sufficiently by better compensation that it is still possible to span the desired distance. On the results of the measurement we actually have to go back to using binary line coding beyond 400 MHz.

The new 6/6A/7/7A cables push this limit ever higher. Thus, with a 7/7A cable we can use a frequency range of 1000 MHz or 1200 MHz with compressed line coding.

Summarising this preliminary study, we come up with the following:

- It is possible to transmit 40 Gigabit/s over STP Cat 7 or 7A

- A transmission band width of 1000 - 1200 MHz is needed

- For 40 GBASE-T the near end cross-talk cancellation must be reduced by approx. 20-30 dB compared with Cat. 6A (for 10 GbE)

- For 40 GBASE-T the far end cross-talk cancellation must be reduced by 20 dB or more compared with Cat. 6A

- The launch power should be between 10 and 12 dBm

- There is no enhanced requirement for return loss. The 8 dB already achieved with 6A is sufficient

Shannon's laws set no limits, things merely get rather more complicated.

On the design of the transceiver circuits

Meeting the requirements of the limit conditions for transmitting 40 Gigabit over STP needs a circuit design that implements modern signal processing algorithms with sufficient bandwidth in circuits with a low power consumption. To do this we need a highly parallel design optimised for the purpose. Although that is extremely interesting in itself, I am only able to give a superficial overview of it for the purposes of these accounts. There are currently small firms such as Solar Flare engaged in designing suitable circuits. In doing so they must work from very unfavourable assumptions, so that a production circuit keeps on working afterwards. On the other hand, looking back on the development of the xDSL circuits, there is experience of this. Here too things have recently simply been tried out in order to overcome the still accepted transmission capacity of telephone cable in the "last mile" – successfully as we see.

The possible decision is to use Pulse Amplitude Modulation (PAM). This form of modulation has proven itself with xDSL and gives an adequate degree of design freedom without setting limitations on e.g. the amount of information that can be coded in one signal transition in stone at the outset. Furthermore, people would like to achieve 40 Gigabit by systematic further development of 10 GBASE-T. There are several reasons for this. 10 GBASE-T is proven and exceptionally cost effective. Users normally already have a structure with 10 GBASE-T running on it when they are considering 40 GBASE-T. In the old tradition the aim will be to develop not chips that can only handle 10 GBASE-T, but auto-sensing 1/10/40 GBASE-T solutions, since these greatly increase volumes and allow the user soft migration. The requirements that 40 GBASE-T makes of the transmission characteristics of the cabling and its environment are relatively stringent. In the course of developing suitable chips the requirements might be evened out or relaxed somewhat.

In what follows we will describe a possible solution for transmission. First of all this involves showing that transmission is possible at all. During a standardisation phase other versions of transmission will emerge and be discussed. Firms like Intel will not allow smaller manufacturers to take the gilt off their gingerbread. With 10 Gigabit over fibre, what actually happened was that Intel bought the small firm with the best technology. This time too there will be a phase during which small developers will run a race whose winner would like to be bought.

Looking at the interrelations described, a bandwidth of 1000 MHz on the cable appears pretty much ideal. Beyond 1000 MHz certain effects are relatively pernicious and the expense of compensation would be high, if not downright impossible.

A practical proposed implementation might provide a Baud rate of 833 MHz, so 833 million signal transitions per second. For 40 Gigabit we then need 48 bits per Baud, which is a lot. On the other hand, we have four pairs of cables, which reduces it to 12 bits per Baud per pair of cables.
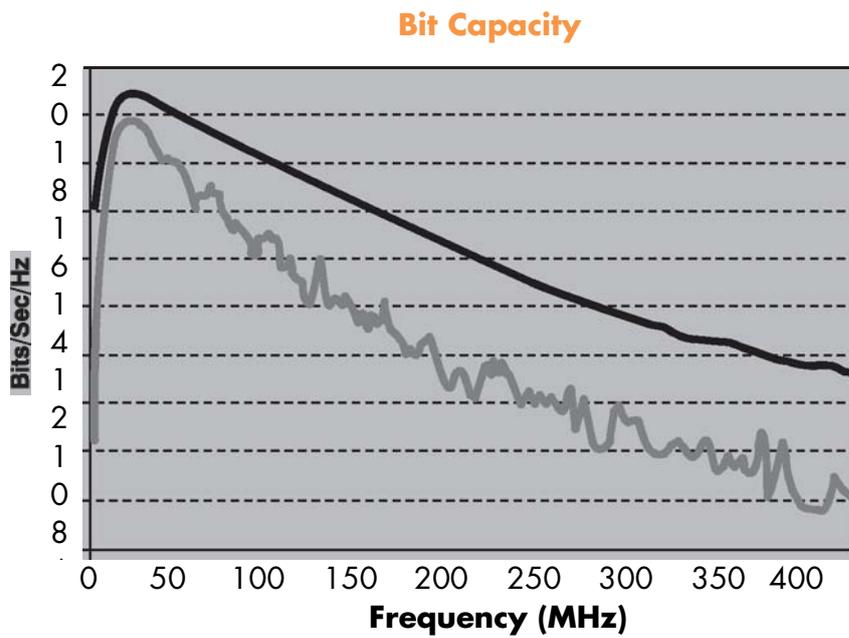
Bit Capacity

Fig 7: Bit pack rate v. Frequency

10 GBASE-T uses PAM-16 modulation. PAM-16 defines sixteen different checkerboard points as a combination of amplitude and phase and hence is able to code 4 bits per signal transition. This is not a particularly demanding modulation technique: for comparison, say, IEEE 802.11a WLANs use up to 64 chequerboard points in the QAM-64. However, one goal of standardisation is to improve the bit error rate. Many Ethernet standards still work on the old default values of the standard, which merely demands 10 EXP -8. However this is no longer in keeping with the times and the standard for 10 Gigabit Ethernet on fibre defines PHY versions that take it to at least 10 EXP -12. It would be desirable to go along with that here. With PAM-16 the general requirements of IEEE 802.3 would be satisfied. However, a clean combination of control signals and Trellis coding needs a PAM-16 that actually uses ten chequerboard points per signal transition. Trellis coding is very complicated mathematically, but we have seen with 1000 BASE-T how well it works and that it appears to be relatively simple to build suitable circuits. For 10 GBASE-T the hitherto three dimensional Trellis coding is extended to four dimensions, each of which is sent over a pair of cables. It is not possible to explain the Trellis dimensions clearly, just like that, but I will try. In two dimensional Trellis coding a surface appears that, so far as I am concerned, may be imagined as a mesh screen. Every cell of the screen can be addressed as standard matrix using two coordinates. Depending on the desired packing of the coding, more or less of the cells of the screen are taken up. Obviously these are located some distance apart. Hence a field appears round every cell of the screen and if, for example, a coordinate is corrupted in transmission, the originally "intended" signal can be recovered. This is a systematic use of the hamming distance in simple linear coding. We now cunningly choose the fields so that adjoining data elements do not fall on adjoining code symbols: rather we place them as far apart as possible. In this way it is possible to achieve much greater security against errors. Three dimensional Trellis coding applies the procedure just described to a solid, which I liken to a dice, on which the code symbols are accommodated equally cunningly. And for four dimensional Trellis coding we take a four dimensional space, which can perfectly well be described mathematically. The idea of using four dimensional Trellis coding and then mapping the four dimensions onto the pairs of cables might almost be described as one of genius, because it results in a gain in the modulation process that immediately boils down to our gaining 6 dB compared with unencoded PAM-16. In this context 6 dB is a lot. The four dimensional Trellis coding means that we gain this 6 dB and at the same time increase the error rate to, say, 10 EXP -12. Incidentally, in order to achieve this error rate, we need a signal/noise ratio of approx. 26 dB on the transmission path for the Trellis coding.

So how do we get any further with 40 GBASE-T? With the ideas sketched out thus far, at least 12 Bits must be represented in one Baud. However, with normal PAM or QAM, we get a maximum of 8 bits to a Baud using QAM-256.

However, QAM-256 is very susceptible to interference and thus unsuitable. Instead of it, we must introduce another intermediate stage in the coding. The information in the 12 bits per signal transition must be represented in a higher order coding, e.g. in a ternary coding. 12 bits gives 4096 different possible states. 8 ternary symbols give 6561 different possible states, so that we would also have something in reserve. A simple line coding for ternary symbols would combine phase and amplitude, so we could take straight PAM-64 of the purpose. However, PAM-64 on the line is not really ideal, since the individual signal levels are very close together.

All things considered, it is better to use OFDM, with which we are familiar from WLANs. The signal is extraordinarily stable. We can use the PAM-64 chequerboard points to modulate the subcarrier directly. The subsequent iFFT-based signal synthesis is known to contribute to a signal that has very low susceptibility to interference. OFDM is in no sense confined to use in wireless networks. In DWDM systems, 40 Gb/sec per DWDM channel is the current basic rate. Discussions on using OFDM for the move to a basic rate of 100 Mb/sec per DWDM channel have been going on for a long time.

A 40 GBASE-T transceiver results from continuing down the old paths that were also successful with 10 GBASE-T:

- Deferral of analogue signal processing
- Improved coding
  - Traditional
    - Gallagher LDPC
    - Co-set partitioning
    - Tomlinson-Harashima
    - 16 or 64 PAM
  - revolutionary
    - QAM instead of PAM
    - OFDM
- Oversampling

There is, in particular, a series of other opportunities for improving the coding, even without OFDM. Gallagher´s Low Density Parity Check LDPC Block Code, for instance, achieves a substantial decrease in the BER as a function of the SNR as concatenated convolutional code (turbo code). 12 dB co-set partitioning provides improved tolerance ofnoise. Gigabit Ethernet uses 6 dB co-set partitioning. 1000 BASE-T 4DPAM-5 transmits 5 levels and thus is as insensitive as 3 levels. 12 dB CSP 4 DPAM-8 transmits 8 levels and is as insensitive as 2 levels. Finally, Tomlinson-Harashima pre-coding provides a reduction in the complexity of the receiver. It allows spectral signal conditioning in the transmitter to reduce the effects of coupling alien crosstalk and eliminates propagation of errors in the Decision Feedback Equalizer (DFE), even at high DFE coefficients.

Another area for problem solving is oversampling in general. There are three alternatives for Transmit Front-End: Simple, Baseline, Oversampled. With "Simple" there is no digital filtering, 3200 Msymbol/sec., simple R/C signal smoothing, the send signal is dependent on imprecise analogue components and there are no spectral nulls at DC and 1/2T, which means poor feedback cancellation. "Baseline" is like "Simple", but with an RLC Frontline Filter with constant output impedance, still with no proper spectral nulls, but with considerably better feedback cancellation. "Oversampled" has digital filtering and interpolation, 6400 Msymbol/sec., simple RC signal smoothing with a base frequency of 1 GHz, well-defined nulls at DC and 1/2 T, and very good feedback cancellation. It is also possible to train PMA sequences by analogy with training OFDM symbols in high-speed wireless.

All these things can be combined in order to achieve the 40 GbE distance target of 100m even with Cat. 7/7A cabling. The job of standardisation is merely to develop the best solutions in terms of economy, power consumption and stability.

Given these preliminary considerations, the requirements for the transmission path can be reformulated: for an aggregate signal/noise ratio of e.g. 25 - 26 dB over the link, the five individual SNR primary factors must have values around 32 dB. These primary factors are:

- Imbalance between channels
- Inter-symbol interference (ISI)
- Echo
- Near end cross-talk cancellation (NEXT)
- Far end cross-talk cancellation (FEXT)

The first two of these you get under control only by a combination of feedforward- and feedback-equalising. For example, a defined sequence of symbols is sent and the result observed. Physically speaking, every existing cable pair in a cabling system has a different bandwidth and, if necessary, a different signal delay. The bandwidth exercises pressure on the inter-symbol interference and the signal delay distorts the 4 X PAM-64 or OFDM useful signal, which is rendered coherent in four dimensions by the Trellis coding. Synchronisation must therefore occur. This problem has been known for as long as Ethernet has existed, even the basic version

10 BASE 5 had to synchronise the receivers to the incoming signal. That is the reason for the existence of the preamble in an Ethernet package. Since traditional Ethernet packages can still be sent in 10 GBASE-T, there is e.g. room available in the preamble for purposes of that kind. Even with maximum-length packages of approx. 1500 bytes, there is the 8 byte preamble, so 0.5 % of the total bandwidth. That is sufficient in any event.

Because of the limited bandwidth, full duplex operation is needed. This means the echo compensation can be performed as part of the direction separation. Mind you, this will take a lot of effort since, because of possible mismatches of impedance, it is necessary to aim for cancellation in the 40 - 50 dB range.

Near end cross-talk cancellation is high-level interference between neighbouring receivers. No matter how much trouble you take with the Trellis coding, distortion takes up a great deal of effort. Here too, you should aim for 40 dB distortion performance for safety's sake.

In 1000 BASE-T far end cross-talk cancellation is described by means of an equalisation parameter (ELFEXT), which is not, however, further compensated by the circuits themselves. Beyond 10 GBASE-T and particularly with 40 GBASE-T, this cannot be allowed and at least one FEXT cancellation must be aimed at the 20 dB range.

As already stated, meeting these requirements, which lead to an SNR of approx. 26, needs a complex circuit design with massive use of parallel structures. One must not forget that part of the work does consist of analogue signal processing, which – in contrast to purely digital processing - cannot be integrated without great care.

In order to achieve the performance demanded, you have to build a MIMO (Multiple Input Multiple Output) circuit that treats all incoming streams of signals in the same way as a matrix filter. This matrix filter has a series of advantages. For example, the near end cross talk cancellation is a process the keeps in turning up in the same way. One pair of cables will interfere with the other cable pair in substantially the same way, since the interference arises from the same output signal. The near end cross talk cancellation for NEXT 1, 2, NEXT 1, 3 and NEXT 1,4, i.e. the three major influences, which are exercised by cable pair 1 on cable pairs 2, 3, and 4, is very similar, but, for example, in the 1000 BASE-T design, is carried out at three different points local to the cable pairs to which the interference applies. This operation can usefully be conflated. With a suitable correlation function, the interference between channels can be cut further. With 1000 BASE-T we always pretend that every signal is surprising and completely new at all times. Obviously this is not correct, since we know the signal very well, because we generate it ourselves in the circuit. A suitably equipped circuit could, for example, generate signals that that are obtained right from the outset in such a way that they cause less interference in a particular environment than an unprocessed signal.

However, these correlation functions can only be implemented if we have the entire signal present in a circuit and not, as hitherto, distributed between four separate, not mutually connected circuits.

It should be clear that it is not enough to carry out the correlations on the sender page alone.

The transmitter must exhibit a linearity of over 50 dB, as must the hybrid-construction receiver. Overall an operational clock speed of approx. 833 MHz is needed.

This means that, as hitherto, the circuit can be implemented using conventional CMOS technology.

Particular thought must go into screening. If you are developing a system for UTP cable, you cannot simply transfer this to screened environments. The screening itself leads to greater interference because of reflections off the screen. This must be dealt with by compensating for the basic background noise in the filters. On the other hand, the reflected signal corresponds to the signal on the pairs of conductors with an extremely low mismatch in time and should be capable of being filtered out by echo compensation.

Another few words about cost. As always in the history of Ethernet standardisation, people would like ten times the performance for around three times the price. This has been proved in the market and has certainly always worked in the past. However, with 40 and 100 GbE this mechanism is disabled. The aim is to have the n-fold performance at the n-fold price or less. Today a 10 GBASE-T board costs approx. 300 Euro.

This means that people would feel that 1200 Euro was an appropriate price for a 40 GBASE-T board. However, from experience the first 40 GBASE-T boards will cost eight to nine times as much as their 10-Gigabit brothers and will then fall in price. In plaintext this means that the first 10 GBASE-T boards will come out at around 1000 US$ and then fall below 500 US$. Hence they will not initially be any cheaper than the equivalent multimode boards, but overall you obviously have to realise that new installation of fibre cabling to e.g. a couple of servers can be considerably more expensive in total than appropriate twisted pair cabling. Obviously this applies in particular if there is already structured STP cabling that can be used. Overall one should initially expect 40 GBASE-T on the server to cost around 40% of an adequate fibre solution: in the course of time these costs will experience a relative fall to approx. 10%. Obviously this generally applies only to links with a maximum length of 100 m and in comparison between copper and multimode.

Overall, the 40 GBASE-T solutions will benefit from the general development process in integrated circuits, since well-known standard processes can be used, more transceivers can be integrated together and there is a considerable potential market.

As we have seen in the last two years, the development of 10 Gigabit Ethernet has indeed proceeded rapidly; however, there were a number of technological problems. As already stated several times, with 40 GBASE even preproduction prototypes in the optical range substantially exceed the requirements of the standards. Only the threat of 40 GBASE-T will cause manufactures of optical components to exercise more pressure on the Tube and on prices. That is exactly how it was with Gigabit Ethernet. Although all the 5-year planning gurus I know are tearing their hair out about this, I recommend implementing 40 Gigabit Ethernet relatively spontaneously when it is needed. The need for 40 Gigabit Ethernet is as sudden and unexpected as Christmas Eve. But we know exactly what happens: the hard nosed get the best prices. If anyone buys Christmas presents in November, it's his own fault. On the morning of the 24th, things are mostly cheaper. And you get the best value gifts after the 26th. Although it looks confusing at first, there are not so many alternatives for the 40 GbE-PHY for an application. If you rush to buy the components, you are in any case clutching at a falling stone. I quoted target prices above. The market will only stabilise to some extent when we arrive at these prices. Anyone who needs a 40 Gigabit server connection now cannot wait for 40 GBASE-T, but will simply take a good value multimode solution. Anyone who now wants to build a metropolitan network can't start anything with a copper solution anyway, but anyone who is newly planning the infrastructure of a computer centre needs to assume that all devices will be connected to one another by monomode.

Consequences for Corporate Networks

As usual IEEE 802 will ensure that the conventional fibre versions are given a head start, so that they will be sold. This can be controlled quite simply via the Project Authorisation Request. The standardisation process as such can proceed quickly – so we will be seeing the first 40 GBASE-T boards in 2011/2012. This is unequivocally within the life of today's newly planned cabling solutions – so these too must be set up accordingly.

With absolute certainty ordinary Cat. 7A cabling will be sufficient for 40 GbE. It is still possible to argue about the connector, but there is no apparent reason why both the GG45 and Siemon should not function adequately.

Another interesting question is obviously what to do about Cat 7 cabling of which there are already multiple installations. This boils down to whether the cabling only just meets the specification or whether, because of the manufacturer's design, there is adequate spare performance. I have intentionally aligned the entire account of transceiver technology with a clock speed of 833 MHz, but obviously higher rates can be used. However, a rate in this range is an ideal match for the existing VLSI process. The use of higher rates may lead to problems with this design process, but this is not known exactly as the process sometimes takes a leap. A clock speed of 833 MHz would, however, make it

possible for Cat.7 cabling to be sufficient for 40 GBASE-T.
Cat 6A cabling is definitely inadequate.

To summarise generally:

- You too will be getting 40 GBASE sooner or later

- The later, the cheaper

- 40 GBASE-LR4 and 40 GBASE-SR just have to be produced by a well-known manufacturer: the components are there and it will happen in 2009 at the latest.

- The situation with switches depends on their basic performance, e.g. a Cisco 6500 can support a maximum of. 2 40 GBASE adapters, so there may be a tie-up with that manufacturer. On the other hand, the 15.000 series can already handle 96 X 40 Gb.

- 40 GBASE-LRM is technically possible at once (AMCC), but is not standardised. This may happen in 2009/10.

- 40 GBASE-T will be included in the standardisation from 2009. There are already suitable cables and connectors, so this will happen quickly.