The Essex Echo: Audio According to Hawksford, Pt. 2

Jan Didden continues his discussion about audio technology with Professor Malcolm Hawksford.

JD: Let's move to amplifier electronics, because one thing that comes across clearly from your publications is that you enjoy electronic circuit design.

MH: Is it that clear? But it is true. My first amplifiers were tube-based, of course, and I still have a certain fondness for them. Most were simple, first-order circuits, with some pleasant coloration usually added by self-induced microphonics and vibrations. Different manufacturers using different tubes even with similar circuits show up different issues, but they err benignly, so to speak. It is very seldom that a tube amplifier's sound can't be enjoyed despite its technical limitations; the errors tend to be quite musical.

JD: What triggered your interest in error correction (EC)?

MH: Peter Walker's *Current Dumping* concept did that. I thought it an extremely clever and elegant solution (still do), and a "thinking out of the box" amplifier design that was *en vogue* at the time. There are various ways of looking at Current Dumping, but I explained it as a combination of feedback and feedforward techniques. The clever bit, as I saw it, was that it allowed you to design a structure that didn't require infinite gain to obtain theoretically zero distortion over a fairly broad bandwidth. In a feedback amplifier, as you move up in frequency, the feedback decreases leading to increasing distortion.

In this (then) new concept, the feedforward path compensates for the loss of feedback with frequency, and in theory you can keep up the "zero distortion" over the audio band. Of course, it depends on what stage of the amplifier produces distortion. It started me thinking about some way to generalize the concept of combining feedforward (ff) and feedback (fb)—which, of course, is at the core of Current Dumping—and explore other trade-offs in ff and fb. As the most objectionable distortion in a power amplifier is generated in the output stage, would it be possible to locally correct that output stage so that the remaining distortion signals that are fed back from the output to the input stage would be much cleaner (i.e., devoid of output stage distortion) thus also contributing to lower input-stage distortion? As N (the uncorrected output stage gain) approximates to 1, the error tends to zero and this makes the difference (correction) amplifier much more linear as it only amplifies small signals, and this holds even when the output voltage swing is large.

The conceptual view (**Fig. 8**) made it clear that, in theory, combining ff and fb can completely eliminate the forward loop nonlinearity, without the need for infinite loop gain, simply by choosing suitable combinations of transfer functions *a* and *b* in **Fig. 8** providing (a + b) = 1. Practical ff or fb networks will most probably need to have some active components and will thus be at least first-order low-pass circuits. But, if the "*a*" network has a first order 1/(1 + sT) characteristic, you could make "*b*" a conjugate sT/(1 + sT), and the elimination of distortion independent of frequency still holds.

Now, for the feedforward component "*b*," there is the practical problem of combining the forward and feedforward signal in the output (power) stage, so that is less attractive. Therefore, one solution would be to use only the "*a*" fb path, as it is much easier to combine low-level signals at the amplifier input. Because you now can no longer compensate for the first-order rolloff, the full curative properties of the system break down at higher frequencies so zero distortion is out of reach.

Yet, employing this type of error correction locally in, for instance, output stages still has significant advantages. Such fast-acting local correction does a good job to linearize the output stage by one or two orders of magnitude and, as a bonus, give very low output impedance before global feedback is applied. I also showed that you can implement a correction circuit virtually without needing more components than those used for biasing, so it's essentially free.

The local loop does not impact stability much, so you can have your cake and eat it, too. You end up with a more linear power amplifier for the same parts investment

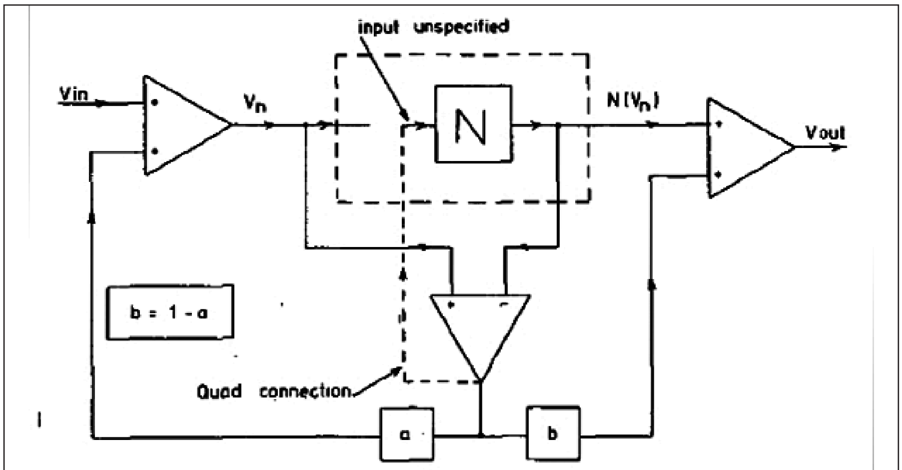

FIGURE 8: Generalized ff-fb error correction structure.

and that's always worthwhile. Bob Cordell had a very elegant implementation of this concept which I like very much[8].

JD: At one point there was a great discussion on diyaudio.com between Bob Cordell, yours truly, and other very smart circuit designers. The question was whether error correction is really a different circuit concept or whether it is another way of using negative feedback (nfb). That it was, to paraphrase evolutionary biologists, a matter of exploring the "space of all possible nfb implementations."

MH: Well, I guess that conceptually it is indeed a different way to apply nfb, but with some interesting different issues which also lead to more insight into this type of circuit. For instance, in **Fig. 8**, assuming that b = 0, then Vout/Vin = G = N/ (aN - (a - 1)). The target for Vout/Vin = 1, so now you can calculate the error function ε representing the overall input-to-output transfer function error, that is the deviation from "1," thus ε is defined as ε = 1 – G.

Substitution gives you ε = (a - 1)(N - 1)/(aN - (a - 1)). Now you immediately see that the error function has two zeros, i.e., (N - 1) and the balance condition represented by (a - 1). This succinctly explains the operation and power of EC, especially with near unity-gain output stages as you get two multiplicative terms in the error function which should both be close to zero. Half of the art of understanding and developing circuits lays in finding the right viewpoint!

JD: I know of at least one commercial implementation of what appears to be your EC concept, based directly on Bob Cordell's circuits, by Halcro. Presumably based on a patent by Candy, which came later in time than your publication.

MH: Yes, I am aware of that. At the time I sent Halcro my papers and wrote to them asking for some clarification, but never received a reply. So it goes. Anyway, life's too short to worry about such things. It's not my problem. Bob Stuart of Meridian Audio also used the circuit in his amplifier range for a period of time, which was most gratifying as he is a very gifted audio circuit and system designer.

There's analogy to error correction in the digital domain, and that is noise shaping. I wrote a paper with John Vanderkooy comparing digital noise shaping with nested differential feedback in analog circuits[9] and concluding that they can be seen as different views of similar issues! If you look at a first-order noise shaping configuration (**Fig. 9**), you see that, similar to EC, you take the difference between the forward block (the quantizer) input and output, which is the noise it generates, and feed it back to the input, properly shaped like H = $e^{(-sT)}$. Now, if you look at the noise shaping transfer function (1 - H), it looks very similar to the error reduction function of EC you showed before. So as you go lower in frequency, where the loop gain gets higher, the noise also gets lower.

Now this is a simple first-order case, but as you go to higher order noise shapers, your in-band noise gets lower at the expense of forcing more and more noise above the audio band. Now, if you put in a coefficient in (1 - H) of less than 1, then the reduction curve bottoms out at lower frequencies, so it is analogous to the bottoming out of your EC curve due to a less than 1 error-feedback coefficient. So, you could say that quantization noise shaping in sampled data systems is analogous to distortion-shaping in feedback or error correction in continuous signal systems. You often see that when the distortion is driven down by feedback or EC, it works for the first few harmonics at the expense of increasing higher harmonic components. Again, just like what we observe with noise shaping in digital systems!

You should look into the literature about Super-Bit Mapping (SBM). Michael Gerzon and Peter Craven in the UK worked on that as did Stanley Lipshitz and John Vanderkooy and also SONY. I well remember a rather heated argument between Michael and a Sony engineer during an AES convention some years ago! The idea with SBM is to apply noise shaping to a digital signal in the context of CD. Normally, with uniformly quantized and dithered 16-bit/44.1kHz LPCM, the

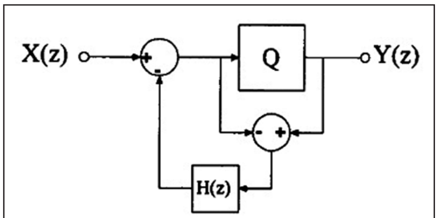

FIGURE 9: Generalized noise-shaping structure.

noise floor is essentially flat from DC to 22.05kHz.

Now, they asked, suppose we start with a 20 or 24-bit source, and we re-quantize and noise-shape the signal, can we somehow retain some of those additional bits of resolution below those 16 bits? Of course, the noise that you reduce in one part of the spectrum needs to go somewhere, and what SBM does is to decrease the noise in the mid band so you get perhaps 18-bit resolution in the frequency region where the ear is most sensitive.

The noise-shaping transfer function is designed to follow closely the Fletcher-Munson curves; consequently, the noise may rise by perhaps as much as 40dB at the very high frequencies, but because your ears are very insensitive in that area you cannot hear it. It is also important to realize that in a properly designed SBM system the noise is of constant level, and there should be no intermodulation with the signal. Also, the signal-transfer function is constant. So, provided that your DAC has at least 18-bit accuracy, you can perceive a subjective resolution of around 18 bit. And at its core, again, is a concept that you would recognize as an error-correction amplifier!

Your use of that AD844 current conveyor in your error-correction amplifier does remind me of a similar topology that I developed with two of my research students, Paul Mills and Richard Bews. This design, which led to the LFD moving-coil preamp, was published in *HiFi News* in May 1988. Richard subsequently developed this conceptual LFD pre-pre that used floating power supply circuitry by optimizing component selection and overall construction to achieve a very high level of performance. The reasoning behind the circuit is as follows: In a simple, single-ended emitter follower (**Fig. 11A**) the transconductance of the stage $G_m = 1/(r_e + R_E)$ where r_e is the intrinsic base resistance.

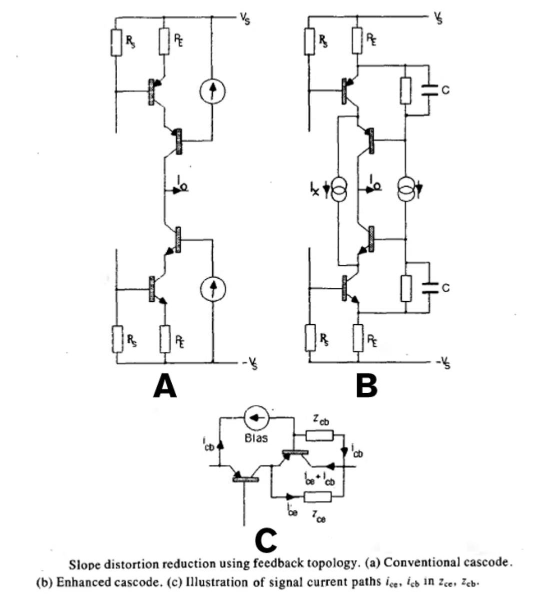

FIGURE 10: Enhanced cascode concept.

Since $r_e = 25/I_E$, you see that because r_e changes with signal current, this introduces distortion. You can improve on this (**Fig. 11B**), and now $G_m = 1/\ (r_{e1} + r_{e2} + R_E)$, where, for example, when r_{e1} increases, r_{e2} falls. There is not perfect cancellation because the transistors of the long-tail pair are effectively connected in series in the AC-equivalent circuit, but it is much more linear than the previous case. You can further improve on that with **Fig. 11C**, where complementary transistors are now effectively in parallel for AC, so the changes in the respective r_es due to signal current are almost perfectly complementary such that the transconductance of the combined transistors is almost independent of signal current; that is, the circuit is linear.

If you plot the nonlinearity (as an error function) versus the value of R_E and signal current (**Fig. 12**), you see that there is a point, with very low R_E, where the **Fig. 11C** stage is almost perfectly linear. So this is a valuable property, but as you can see there are some challenges in biasing it, especially with those very low-value emitter resistors. However, you can rework the circuit to retain the linearity yet make biasing somewhat easier.

Another most important aspect of the topology is the use of truly floating power supplies because even if the supply voltage were to vary or to exhibit noise, there is no signal path linking to the RIAA impedance, as related currents can only circulate in closed loops. Consequently, power-supply imperfections are dramatically reduced, which is very critical in MC applications where small signals can be sub microvolt in level.

Under large signal conditions, you have transistor slope resistances and slope capacitances which are being modulated by the signal, and that's potentially bad news. Some people call it phase modulation, going back to something Otala brought up many years ago.

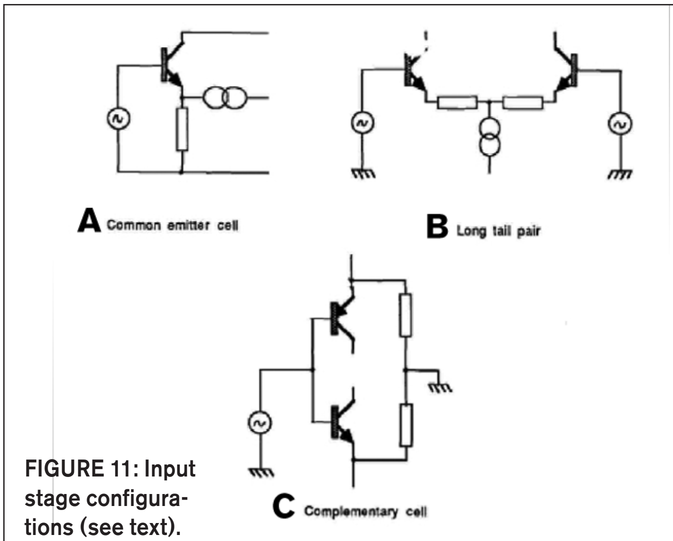

FIGURE 11: Input stage configurations (see text).

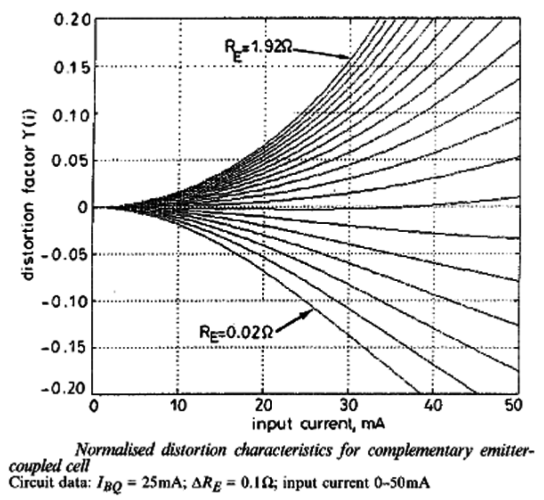

FIGURE 12: Input stage linearity versus RE.

It's more like a gain-bandwidth modulation, and I prefer to think of it as a time-domain modulation. For instance, in a feedback amplifier, this would slightly modulate the open-loop gain-bandwidth product and you can then calculate what it does to the closed-loop phase shift. It's like a signal-dependent phase shift, which manifests itself as jitter. It is analogous to a signal-dependent jitter, and it basically happens in all analog amplifiers. So, you have jitter in digital systems, you have jitter in I/V converters due to finite slew rate leading to slight modulation of the loop gain-bandwidth product, and you have these signal-dependent jitter-like phenomena in analog amplifiers in general, albeit that the modulation is time continuous rather than being instigated at discrete instants.

PHOTO 3: A younger Malcolm Hawksford showing off his speaker building skills.

You know, if you start to design a system, you need to have some sort of philosophy that drives you. For me, it is often the minimization of these timing errors, and I think that large-signal nonlinearity is less of a big deal than sometimes is believed. Most of the time you listen to low-level signals anyway, where linearity is very good. So then, you ask, what distinguishes one system from another, right in these low-level regions?

Now, I don't have any magic number, but let's assume that 100pS is the magic number for digital jitter, and suppose that you find similar numbers for what I call "dynamic timing errors" in analog amplifiers, the picture sort of comes together. It just *might* be that simple, open-loop circuits, while having higher large-signal level distortion, potentially have less of these timing nonlinearities, which could explain their very good sound. I would need to get the sums together, but it just might be possible that this is one of the reasons why people prefer those simple, low-feedback amplifiers. Especially in transistor circuits, where the transistor parameters themselves are modulated by changing voltage and current.

So having simple circuits that minimize these changes and are designed to minimize power supply influences clearly helps. Of course, feedback can help in many ways and there is no fundamental reason that a feedback amplifier cannot exhibit exemplary results, providing care is taken to minimize modulation of the amplifier loop transfer function.

Another example: When Paul Mills was still at Essex, he was working on an amplifier design using a cascode stage (**Fig. 10A**) that had reasonably low distortion. Then I told him, "Look, Paul, I will make one modification to your circuit that lowers the nonlinearity by an order of magnitude!" What I did was re-locate the biasing for the cascode to its emitter rather than to the supply (**Fig. 10B**).

It doesn't look like much, but it is a very significant change, and I can explain it with **Fig. 10C**. Why is the Z_{out} of a cascode not infinitely high and its distortion zero? It has to do with transistor slope parameters and their modulation with signal level.

PHOTO 4: Professor Hawksford and PhD student Adam Hill in the university audio lab.

You can see that an error current that is the difference between the ideal output (collector) current and the actual one is a result of the non-infinite impedances between emitter-collector and base-collector of the cascode transistor, where in **Fig. 10C** these two impedances are modeled by Z_{ce} and Z_{cb}. The modulation of transistor slope parameters with signal level I mentioned can be described as modulation of Z_{ce} and Z_{cb}. So, if you could find a way to prevent these error currents from ending up in the output (collector) current, then their bad influence would be eliminated.

Now, what is the effect of re-locating the bias to the emitter instead of the supply? For example, the i_{cb} error current now no longer comes from the supply but from the emitter of the top transistor. It is subtracted from its emitter current, which is basically the same as the cascode collector output current. So when i_{cb} is added to the cascode output current, it is no longer an error but makes up for the current that was subtracted in the first place! For i_{ce} a similar reasoning can be made. So the error currents now circulate locally in the stage and don't contribute to the output. It doesn't work perfectly, because there are some minor errors due to base currents, but it is, nevertheless, a huge improvement. The output impedance goes up typically by a factor of 10, and the distortion goes down by a factor of 10!

Note that it does not matter whether these error currents have a nonlinear relationship to the signal, as they do not contribute to the output current. This technique therefore works well in large signal amplifiers. I just picture this process in my mind, and I "see" what's going on, and then the solution pops up.

JD: You need to make the mental leap to model this modulation as an error current, and then find a way to shunt that error current away.

MH: Yes, indeed. There are some issues involving stability, as there is some form of regeneration in the circuit, but that's the gist of it. Now, I often wonder whether I would have seen that if I had plugged it into a simulator and run a distortion analysis. I like to think that I might *not* have made that connection. I also believe that you should lay out the PC board, build your designs, and think about the topology at the same time. The days of a light box and black tape were great and very intuitive, very human. You move the layout around, changing this and that and in some way that connects back to the circuit again and you may then end up improving the circuit. It's an iterative process that can give you just that extra bit of quality or performance that you don't get when doing a sim and then saying, well, that's it.

Anyway, this particular enhancement then appeared in my enhanced cascode paper[10]. Also, Richard Bews and I used this concept in the LFD preamp (**Fig. 13**), which, as previously mentioned, employed a true floating power supply system. And even if those batteries were to introduce some supply voltage nonlinearity, this doesn't show up in the output signal. There are no grounding problems because of the floating supplies. The floating-bias input pair is coupled to a cascode stage.

It's clear that any changes in that bias voltage do not have any influence on the output signal. So this will have high output impedance which drives current into the passive RIAA network to convert that current to voltage.

Now, if you look at which components determine the sound quality, it's only the input transistor emitter resistances and the components of the RIAA network. The cascodes don't do anything; the power supplies don't do anything, so it's an extremely linear circuit overall. And because it is only those few components, Richard was able to optimize component selection, ending up with a truly world-class preamp. Richard really is extraordinarily good at tuning and laying out circuits, and the battery-powered preamp worked extremely well. Also, this is why LFD Audio now enjoys almost cult status with its amplifier products.

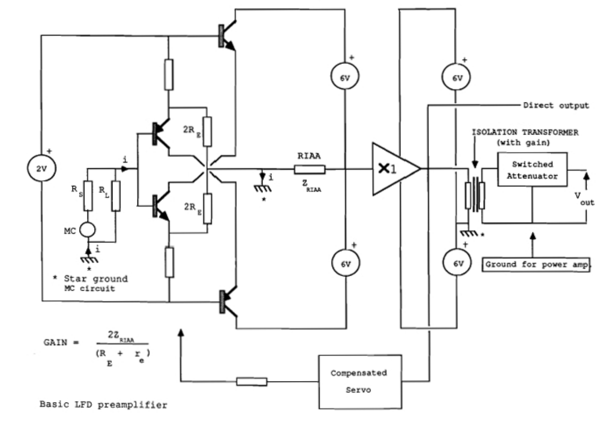

FIGURE 13: LFD preamp simplified diagram.

JD: That **Fig. 13** circuit looks deceptively simple, but it is a very intricate circuit, isn't it?

MH: Yes, it is very simple, yet has a lot of interesting points: low noise, low distortion, almost no supply interaction, virtually no ground-rail current, very insensitive to transistor parameters, accurate RIAA correction, yet only a few active devices.

Often manufacturers have a good basic topology, but then they need to work in the power supply and grounding as well as the electrolytics and the other components in the signal path, and it all tends to blur the final sound. If you have many components in the signal chain, individual optimizations have relatively small impacts. But with this simple circuit, the components that determine the quality are few, and thus optimization has a relative large effect as well.

The absence of power supply interaction, however, is key to its performance. I find that at least as important as the topology itself, not only in preamps, but also in DACs and power amplifiers, for that matter. A lot of the differences between equipment in terms of clarity and cleanliness have to do with internal EMI issues and the power supply interactions and ground contamination.

JD: Well, we've already covered a lot of ground, but perhaps I can ask you about your views of switch-mode amplifiers.

MH: As you know, I've done a lot of work on Sigma-Delta (SD) modulation over the years. There is one proposal using an SD modulator driving an output stage with a pulse-density modulated signal. Now, the switching frequency would generally be higher than in the case of a PWM stage.

As you mentioned before, there is a basic problem with these types of circuits with EMI, and a higher switching frequency doesn't help. Do you remember our discussion with raised-cosine modulation in a DAC? Well, in this particular idea I used something similar. Instead of supplying the switching output stage with a stiff supply, you use a resonant supply synchronized to the switching frequency of the amplifier. The supply voltage would, in effect, be a raised cosine, so that at each switching instant the supply voltage would be zero, and would then smoothly rise toward the full value (**Fig. 14A**).

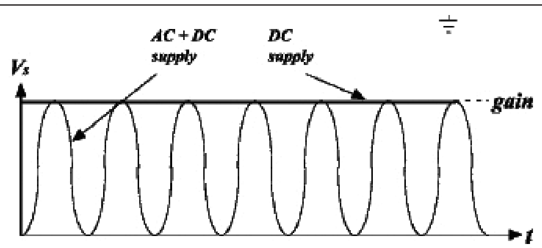

FIGURE 14A: Resonant power supply synchronized to sample rate outputs raised-cosine voltage.

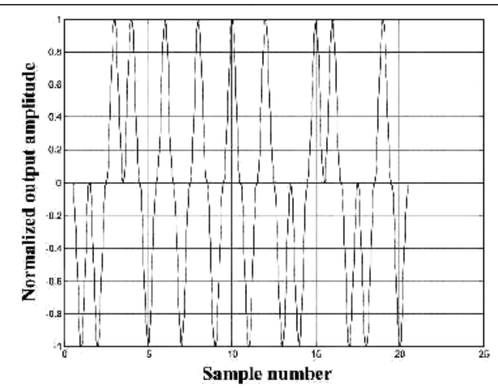

FIGURE 14B: Raised-cosine supply for switching amp dramatically reduces output signal bandwidth.

The result is that EMI problems are greatly reduced because the switching effectively occurs at zero voltage, and the harmonics are both lower in level as well as much lower in bandwidth. The output voltage of the amplifier is now no longer rectangular but somewhat sine-shaped (**Fig. 14B**). Switching efficiency of the output stage is improved as well, and not only are those switches still either fully on or fully off, but because switching occurs with zero voltage across the device, power dissipation in the finite switching transition region is reduced. The average output level of this scheme is somewhat lower than a regular PWM amplifier, but that can be compensated for as described in the paper.

JD: Do you think that these switched-mode amplifiers can reach the quality levels of a good analog amplifier?
MH: Well, I've heard some commercial systems with B&O IcePower modules, which seemed to work really well, so I would say it's getting there, yes. It's an interesting technology, and even if the samples I've listened to were not always very low distortion, they did have a certain cleanliness and transparency to them. I'm not absolutely sure, but it may be related to the absence of low-level analog problems like dynamic modulation of device characteristics in an analog amplifier.

So, I'm fairly optimistic, also because it brings the digital signal closer and closer to the loudspeaker, skipping analog preamps and the like. Of course, you need to distinguish between "analog" switching amplifiers and "digital" switching amplifiers where the power amplifier is, in effect, the DAC. I have always been more interested in the latter class, especially the signal processing needed to achieve good linearity[11,12]. Just because an amplifier uses switching techniques does not necessarily make it a digital amplifier. This is an important distinction which is often misunderstood.

JD: Bruno Putzeys, a well-known designer of switching amplifiers, maintains that switching amps are analog amps: they work with voltage, current, and time—all analog quantities.
MH: Indeed. So, there are still a lot of problems to overcome, but they have a philosophical "rightness" about it.

JD: Not the least because of the high efficiency!
MH: Of course. And even if you want ultimate quality, running your amp in class-A with a 500W idle dissipation doesn't solve your quality issues either. There's much more to amplifier quality than just the choice between class-A or class-AB/B topology. An AB/B amplifier, properly implemented, with attention to all the often misunderstood issues of biasing, power distribution, grounding, and so forth, can sound so good that there is nothing to be gained by going to class-A. It's better to go for a simple system, with as few stages as possible because an additional stage cannot fully undo any damage done by a previous stage.

Now, a great-looking box with lots of dials and lights certainly may play music well, but for ultimate quality, get the best DAC you can afford (preferably a networked DAC linked to a NAS drive!), followed by a passive volume control and a great power amplifier and, of course, keep the cables short. Nothing can beat that, in my opinion.

JD: Professor Hawksford, thank you very much indeed for many hours of your time, for most interesting and illuminating discussions. In particular, I was intrigued by the correspondence between seemingly disparate phenomena, like noise shaping versus error correction and jitter versus analog phase modulation. I hope this will inspire readers to do their own experiments and come up with yet other interesting configurations. *aX*

REFERENCES

8. A MOSFET power amp with error correction, presented at the 72 AES Convention, Anaheim, 23-27 Oct. 1982, revised 25 Jul. and 27 Oct., 1985. Also on www.cordellaudio.com/papers/MOSFET_Power_Amp.pdf

9. Relationships between digital noise shaping and nested differentiating feedback loops, presented at the 93rd AES Convention, San Francisco, 1-4 Oct., 1992, revised 9 Nov. 1999.

10. "Reduction in Transistor Slope Impedance-dependent Distortion in Large Signal amplifiers," *J*AES, Vol. 36, no. 4, April 1988.

11. "Dynamic model-based linearization of quantized pulse-width modulation for applications in DA-converters and digital power amplifier systems," *J*AES, Vol. 40, no. 4, pp. 235-252, April 1992.

12. "Linearization of multi-level, multi-width digital PWM with applications in DA conversion," *J*AES, vol. 43, no. 10, pp. 787-798, October 1995.

Note: All papers referenced here and many more are available at www.essex.ac.uk/csee/research/audio_lab/malcolms_publications.html