

What's Ahead in High-Speed Wireless Data Communications? The Future Will Be Better Tomorrow – And Different Than What We've Been Expecting

Daniel Foty (MIEEE) and **Klaus Runge (MIEEE)
 Gilgamesh Associates, Fletcher, Vermont, USA

&

**RFive Communications, Westlake Village, California, USA

Abstract—The present situation in high-speed wireless data communications is examined. While there is growing demand for wireless bandwidth, the most pressing problem affecting this situation today is the attempt to increase bandwidth by using the same technology with tricks - rather than by using innovation. Opportunities for innovation are quite good with higher carrier frequencies, since these enable simplicity and low power consumption – opening the door to truly portable wireless peer-to-peer (WP2P) networking. Numerous challenges exist in technology and design methods; however, meeting these intellectual challenges is the only route to new and exciting wireless data technologies.

Index Terms—Wireless communications, networking, bandwidth, peer-to-peer networking

I. INTRODUCTION

ESPITE various ups and downs, long-term demand for wireless bandwidth continues to increase. At this time, though, attempts to increase bandwidth are hampered by self-constraints of using present-day technology and a variety of “tricks” – rather than using innovation.

This paper will offer a new and alternative vision for high-speed wireless data communications - in which innovation is a central feature. The use of higher carrier frequencies is the only way to reach higher data rates while preserving system simplicity and low power consumption; simplicity and low power consumption are cornerstone requirements for development of portable devices which can function in a truly wireless peer-to-peer (WP2P) networking infrastructure. This innovative approach presents many very difficult challenges – challenges which are both tangible and intellectual; meeting these challenges will open the door to new capabilities in high-speed wireless data technology.

II. GROWING DEMAND FOR WIRELESS BANDWIDTH

In what follows, we will employ the principle articulated by the great American philosopher Yogi Berra: “You can observe a lot just by watching.”

A. General Drivers for Bandwidth Growth

Over time, the demand for bandwidth (both wired and wireless) continues its relentless increase. Occasionally, as happened with optical communications in the late 1990s, the supply of new bandwidth outruns the short-term demand; this kind of occurrence is sometimes interpreted as indicating an end to the need for more data bandwidth. However, the long-term trend is clear - there is a strong and steady growth in the demand for more bandwidth. This demand resembles the situation in equity markets – there are down periods (bear markets) where the return on investment is not good, yet the long-term trend is for a general appreciation of equity prices.

As capabilities in both wired and wireless data communications improve, the concept of “anytime, anywhere” communications moves closer to becoming a reality. This idea has been discussed for more than a decade (see, for example, [1]), and it has been noted at least that far back [2] that a high-quality wireless data communications capability is required for this vision to become reality. This is illustrated by a 1994 postulation (Figure 1 [2]), showing a variety of layers and capabilities for access to data.

Already, there is an emerging variety of wireless data networks; these are best exemplified by the proliferation of WLAN “hot spots” in a variety of locations, such as cafés and airports. At present, though, these “hot spots” are confined to relatively restricted geographical areas (generally no larger than a typical city block); generalized roaming is not yet a feature of WLAN technology.

In terms of truly-portable data, with network access on a “roaming” basis, capabilities are appearing on the scene and are improving. For example, the

“EDGE” network is appearing more widely in urban areas around the globe; recently [3], for example, it was announced that an EDGE network will be deployed in Moscow. The EDGE network offers data speeds ranging from 130kbps (uncompressed) to 250kbps (compressed). In addition, the long-delayed but slowly-appearing “3G” networks will offer data speeds of 2 – 2.5Mbps. While these offerings represent improvement, they also make it clear that wireless data speeds lag behind WLAN hot-spot data speeds (now 54Mbps) and far behind wired data network speeds. For comparison, in wired protocols, the Ultra3 SCSI speed is 160Mbps (equivalent to 1.28Gbps), while the data rate of the FC-AL Fiber Channel is 100 – 400Mbps (equivalent to 800Mbps – 3.2Gbps). Clearly, wired data speeds continue to far exceed wireless data speeds; this impedes the capabilities of portable devices, and also raises questions about what will happen at wired/wireless interfaces.

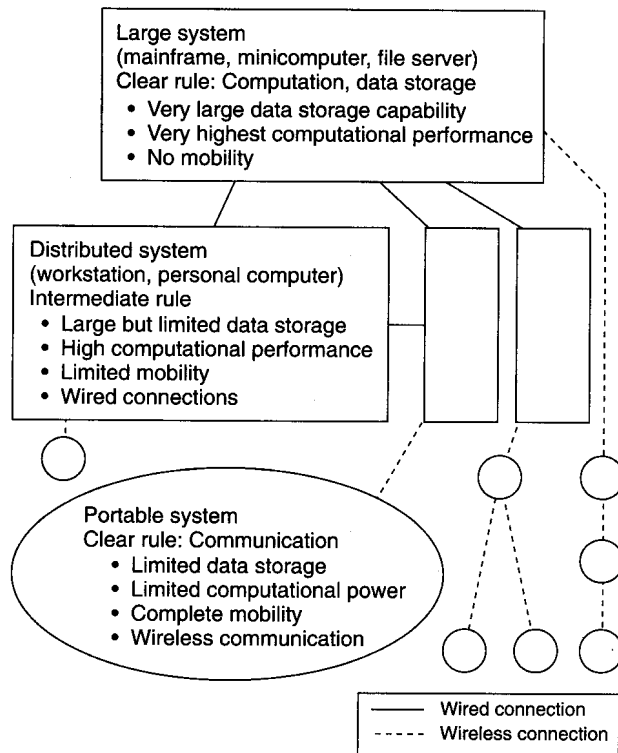


Figure 1. A 1994 vision of an integrated computation/communication infrastructure [2]; this concept is still a work in progress, requiring further advancement in high-speed wireless data technology.

Amidst these infrastructure constraints, demand for wireless bandwidth continues to grow. In particular, there is an increasing use in portable environments of separate and converged forms of audio, video, voice, and data. There is also a growing need for security, which increases the system overhead.

Some generalized drivers of demand for wireless

bandwidth are:

- Connectivity for “appliances” (in offices, homes, cars, etc.)
- Sensor networks (both embedded and “free-range”)
- Wireless data access “as good as in the office”
- Streaming audio and video – “Infolmaging”
- Various forms of “tele-presence” (such as “tele-medicine”)
- Numerous aerospace and military applications.

B. Specific Drivers for Bandwidth Growth

According to InStat, shipments of “internet-access-capable-devices” in 2004 reached some 791 million units, with that number projected to grow to 1 billion units in 2008 [4]. In general, this represents the proliferation of “smart devices” which will communicate and synchronize wirelessly.

More and more devices (both portable and otherwise) are becoming wireless-data-capable; a list of such “devices” includes (but is not limited to) PCs, peripherals, PDAs, mobile phones, digital cameras and camcorders, portable audio/video players, sensor networks, automotive components, and military/aerospace technologies. Another interesting note is that in the United States the Federal Communications Commission (FCC) is pushing forward with a shift toward high-definition television (HDTV) along with a planned phase-out of the old transmission protocol (which will disappear within the next ten years). HDTV has an uncompressed bandwidth equivalent of about 1.5Gbps; even with compression technology, very high wireless data bandwidth will be required for the distribution of HDTV signals around a “space,” such as a home.

There is also a growing convergence of formerly-separate devices into multi-purpose “super-devices.” A prominent example is the rapidly-emerging dominance of “super-phones” which combine together the functions of a mobile phone, a PDA, and even a portable PC; this has actually caused the sale of “pure” PDAs to flatten (with a downward trend projected in the future) as PDA functions are absorbed into mobile phones. InStat [4] also projects that shipments of internet-access-capable mobile phones will reach 591 million units in 2005.

These trends make it clear that portable devices are offering better and more diverse capabilities. As these devices become capable of doing even more, they will require more wireless bandwidth.

III. PROBLEMS IN INCREASING BANDWIDTH

While the demand for wireless bandwidth continues to grow, present-day approaches to improving

wireless bandwidth are encountering limitations. These limitations are self-induced, as technology can only be pushed so far. The various causes of these limitations must be understood before a new vision for addressing these needs can be articulated.

A. General Difficulties

At this time, most wireless data technologies employ relatively low carrier frequencies – 5GHz or less. There are a number of interesting difficulties in this frequency range, largely due to signal attenuation problems; for example, the use of a 5GHz carrier hasn't worked out well, due to strong absorption by commonly-used building materials. As a result, almost all attention has focused on the use of a 2.4GHz carrier frequency; this frequency is in an unlicensed spectrum band, and offers relatively good transmission characteristics in interior environments. Today's varied "WiFi" technologies use a 2.4GHz carrier.

While this band has been very successful, it is in many ways becoming a victim of that success. From a regulatory perspective, the 2.4GHz band is becoming very crowded, as it hosts all of the various 802.11 WLAN technologies (and successors) as well as the Bluetooth short-range data protocol. This band is also used by microwave ovens; this raises several concerns. Obviously, there is a considerable risk of interference with wireless data signals. In addition, there are health and safety issues; as its use by microwave ovens indicates, 2.4GHz is literally a "cooking" frequency; while the technology could accommodate higher transmission power levels, this option is limited because it could literally heat people and cause health problems.

While there is presently an alphanumeric soup of protocols and "standards," these basically all involve the use of a 2.4GHz carrier; the differences among the protocols and standards amount to clerical matters in handling data and modulation schemes. For any given carrier frequency, there are engineering limits; ultimately, it is only possible to push so much data over a particular carrier frequency. There is only so much that can be done by increasing the cleverness of the modulation scheme – such as the increase in the data rate from 802.11b (11Mbps) to 802.11g (54Mbps).

Ultimately, when one takes a "big picture" perspective, it becomes clear that the data rate is merely one of many interlocking factors – factors which trade off *against each other*. The list of such items is quite lengthy, including (but not limited to) data rate, power, transmission distance, form factor... and also cost and flexibility. The key item that must be understood is that under this kind of circumstance, *the data rate can only be increased by paying a price in the other engineering factors*. This is simply a fact

of engineering life – design decisions are ultimately based on trade-offs, and indeed "There is no free lunch."

A common path used in such situations is to resort to "tricks" to try to extend a technology further than it would normally go. This can often provide short-term relief, but rarely opens a way to a long-term result. In terms of wireless data communications, the use of "tricks" causes that technology to become more complex, more power-hungry, more expensive and less flexible – as will be discussed in more detail shortly.

A good example of how aspiration may outrun capability can be found by looking at some of the further alphanumeric soup which is being discussed as part of the "standardization" process [5]:

- 802.11n, with projected data rates of 108 – 520Mbps;
- 802.15.3, with projected data rates of 480Mbps – 1.3Gbps;
- 802.16, "wireless metropolitan-area network" (WMAN), as an alternative to xDSL;
- 802.20, "true mobile broadband."

All of these look great on paper, but a simple question has to be asked. How will these technologies achieve the desired data rates? Will they rely on "engineering tricks"? Or will they be delivered via genuine innovation?

From a simple-minded (but correct) perspective, there are basically three ways to increase the data rate (bandwidth). The most "brute force" method is to simply increase the transmission power; the simplest method is to use a higher carrier frequency; the third is to use a variety of engineering "tricks" to try to extend technology beyond its expected limits. The last option, the use of "tricks," is usually either very clever or very clumsy (and often both at the same time). Some of the proposed "tricks" should be examined to see if they really offer a promising route to higher data rates.

B. Attempts to Extend Low Frequency Technology

In this section, two methods of extending low frequency wireless technology are examined.

One such technology which is currently receiving a great deal of attention is "MIMO" – "multiple-input, multiple-output" wireless data technology. The basic idea behind MIMO is that if one radio isn't fast enough, then the "solution" is to put multiple radios in parallel to increase the overall effective bandwidth. There is a historical echo here; during the 1990s, when dial-up modems reached their speed maximum of about 56kbps and before alternatives (such as DSL and cable modems) became available, there was a brief flurry of "parallel 56kbps" modems. When a site had two separate telephone lines available, these

modems used the two phone lines in parallel in an effort to increase (double) the effective bandwidth. Of course, there was considerable technical overhead, since data had to be split at the source and re-constituted at the destination. These products were clumsy, and were short-lived as simpler and better options became available.

MIMO technology uses the same strategy, and since it does not face wiring constraints it is possible to include several parallel radios. While this may seem clever, it is in fact a rather clumsy approach – to achieve a higher overall data rate, a serious engineering price must be paid.

First, there are serious form factor issues, as MIMO devices require multiple independent antennae for the parallel radios; MIMO hubs presently on the market resemble porcupines with their multiple protruding antennae. If MIMO is deployed in small, truly-portable devices, multiple antennae will have to be swung out during use – forcing portable devices to resemble Swiss Army knives. Beyond the questionable aesthetics of such multi-antennae systems, there are serious increases in the risk of damage or breakage due to all of the protruding antennae.

Second, any MIMO approach intrinsically imposes on itself a *very large* computational overhead, as it is necessary to “break up” data for transmission and reconstitute it on the other side. This computational overhead requires a considerable amount of power consumption - due to the complex modulation schemes required (and some other factors). There are thus additional form factor issues besides just the multiple antennae; the considerable degree of baseband computation requires the use of much more silicon real estate, and the system boxes must also become larger. With this sort of implied demand for more complex engineering, there are worsened problems with cost, reliability, etc.

It must finally be noted that “scalability” in MIMO systems is strangely self-justifying – higher data rates are obtained by “simply” adding more parallel radios. In this approach, the problem is addressed by making the engineering larger rather than by making it more clever. At this point, it is worth noting the observation of the late Czech-American engineering professor and scientific philosopher Petr Beckmann, that “In a healthy society, engineering design gets smarter and smarter; in [an unhealthy society], it gets bigger and bigger” [6]. Viewed in this light, MIMO is ultimately bigger rather than smarter, and seems less an innovation than an act of desperation. MIMO may offer some short-term relief as demand for wireless bandwidth grows – but it is not a viable long-term strategy.

A second salient example of the difficulties inherent in trying to extend present-day technology to higher data rates may be found by examining the recent

proposal for an upgraded wireless USB (WUSB) protocol – one with data rates equal to the USB2.0 rate of 480Mbps. That data rate is achieved by using the engineering trade-offs mentioned earlier – in this case, by giving up quite a bit in both power consumption and in transmission distance. In terms of power consumption, WUSB portable devices aren’t envisioned as being truly “untethered” – they are “portable” when doing “portable things,” but must be in a power-line-connected cradle when doing WUSB things. This kind of approach requires the use of a “star” (or “hub-and-spokes”) configuration, which is naturally used in wired communications topologies. However, in this form, WUSB is not truly a portable technology, and is certainly not capable of wireless peer-to-peer networking. In terms of transmission distance, the data rate for 480Mbps is very short range – going out to only **two meters** and then falling off to 110Mbps at a distance of just ten meters. Thus, devices using this technology must be kept very close together when engaged in wireless data communications. With these constraints, it is clear that the only path for scalability to higher data rates is to reduce the transmission distance and to use more power. The self-constraints on this technology are very severe, and can only become more severe under any attempt to extend that technology.

C. Regulatory and Standards Issues

One structural problem with wireless communications is that, by definition, governmental regulations will be an issue. It has already been decided that governmental organizations (both national and international) should have control over the use of the electromagnetic spectrum for communications. We can only hope that by concerted efforts, these regulatory bodies will be able to adjust laws and rules to reflect present (rather than past) realities.

To add to the *technical* problems, the entire effort in wireless data communications has unfortunately ceded too much of its engineering decision-making to committees which set “standards” and “specifications.” There is a fine line between “standards” and “central planning,” and in many cases this line has been crossed, to the detriment of nearly everyone. We now have over-standardization, under which too much engineering is embodied in (dictated by) “standards.”

A fundamental technical danger is that this situation serves to stifle and limit the scope of innovation. In addition, under such conditions over-standardization distorts the relevant business constraints, and ends up (in the short term) favoring buyers over suppliers. As standards have become too specific and too all-encompassing, there is no “play” left for innovation; as a consequence, everyone ends up designing “Me-too” parts from the outset. From an integrated circuit

viewpoint, this forces designs to begin their product life cycle **already as commodities**. The business dynamics of the semiconductor industry have made themselves clear over a number of years – there is a requirement of a pre-commodity phase in order for a product (and a company) to be profitable, and thus to justify the investments which are required. If this phase is allowed to be “regulated” away, then there will be no profitability for integrated circuit suppliers – suppliers go bankrupt, and risk capital managers sensibly refuse to make any further investments in innovative new start-ups in this area. Eventually, buyers find that there are either no parts to buy - or that those available are of the low-quality “junk-food” variety and remain stagnant over long periods of time.

A good example of this unfolding of events is the ongoing “bloodbath” in the 802.11 chipset market. Recent times have seen numerous bankruptcies and “exiting of the business” moves, as it has proven difficult to run a survivably-profitable business in this market space. Basically, one cannot innovate if one cannot survive. As a consequence of these happenings, investment in 802.11 chip set start-ups has basically evaporated; we are once again learning the lesson that conditions must be friendly to attract the good investment flows which are required to fund successor products.

A final concern with standards and specifications is that “Those who ignore history are destined to repeat it.” It is often forgotten that ten years ago, the RF-IC market was a very difficult one to enter due to the opposite problem – a complete lack of *any* forms of standardization. Instead, the marketplace was badly fragmented and Balkanized into a large number of very tiny segments. Because each segment was so small yet required very specific design objectives, the risk to product development was very high – quite literally, with such a limited marketplace and a very small number of potential customers for any particular product, it was difficult to justify investments (particularly in start-up companies) when success was completely dependent on acceptance from a pool of only two or three potential customers.

The recent proliferation of a very large number of “standards” (in the alphanumeric soup) seems to be dangerously reprising the “bad old days” of such things. In trying to mediate between conflicting user interests, there is a growing risk that fragmentation is returning, with a concomitant return of the problems discussed above. While there is some level of need for standards and protocols, there must be a new focus on keeping the number of such things as small as possible, so as not to fragment the marketplace into uselessness. In particular, more decision-making should be left to the marketplace rather than standards committees.

The key point of this discussion is to note that there is a need in the RF-IC marketplace for a relaxation of the “standardization” process, which has gotten dangerously out of balance. This is particularly important for designers of RF-ICs, who must realize that they can’t simply sit back and wait for someone else to write down all the details of a “standard” for them – this is a road to bankruptcy. Over-standardization also has a tendency to breed unrealistic expectations among design groups; it must always be asked, “If you are going to use the exact same inputs as everyone else, why are you so special that **you in particular** are going to prosper?” Entering a competitive marketplace based solely on the belief of personal superiority is *occasionally* justified – but this is *rarely* the case. The best way to succeed in a competitive marketplace is via sound and effective *innovation*.

IV. AN OPPORTUNITY FOR INNOVATION

As has thus far been described, progress in high-speed wireless data communications is now encountering many severely limiting factors – factors which are technical and structural. While it is a fine thing to put out goals and hopes, the ability to push the present approaches to wireless data communications, already showing signs of strain, is severely self-limited. New thinking and new innovation are required to realize the potential that was noted at the beginning of this paper. Some ideas along these lines will be discussed here.

First and foremost, higher carrier frequencies will be required. Basically, only the use of higher carrier frequencies can maintain (or recover) simplified approaches to communications design at all levels of engineering. This also more easily facilitates the deployment of “intelligence” into *baseband*, where it is very reasonable to create the ability to perform many simple “digital” tasks, such as being able to use multiple data protocols in a single transceiver system.

Second, wireless data communication needs to shift its focus away from the “legacy” use of the “star” (hub-and-spokes) configuration – which is indeed a “legacy” of the methods of wired communications. The dependence on the “star” configuration simplifies the engineering challenges – but it also greatly circumscribes the future potential of high-speed wireless communications. Instead, the focus must shift to wireless peer-to-peer (WP2P) networking among truly untethered portable devices. Viewed in this way, it is made very clear that major innovations are required on two fronts.

Data rates must be greatly increased – they will need to be as good as available wireline data rates. Ultimately, the vision of “anytime, anywhere” communication and computation requires that there

be no meaningful data rate difference between a wired connection and a wireless connection. As a facilitator, wireless data technology must also offer ease of scalability to higher data rates; this must be feasible with minor technical improvements rather than by repeated “from-scratch” technology development

In addition, power consumption must be very “low.” While “low power” is regularly discussed and claimed, some benchmarks can be set for the context under discussion here. “Low power” must mean “sufficiently low” to allow portable devices to be *truly* portable – that is, they must be able to perform *all* of their capabilities and tasks over long periods of time while not connected to line power. This is a requirement if portable devices are to be suitable for truly portable high-speed wireless peer-to-peer networking *among multiple devices*.

Furthermore, major improvements in the power consumption provide a “budget” which can be used to increase the capabilities of the technology. The simplification of the transceiver technology permits increased functionality to be implemented **at baseband** and thus in software – where such things are used to the best advantage. This also enables a wireless device to engage multiple peers simultaneously, since clerical multi-protocol tasks can be handled in a “digital” fashion. In addition, low power consumption allows for “reasonable” transmission distances – “reasonable” being on the order of 100m – as an enabler for real WP2P networking.

By simplifying the transceiver, wireless networking can be done using a single channel and a single antenna, while employing a simple modulation scheme. This, in turn, allows for much more pleasant form factors for modules, equipment boxes, etc.

Of particular note at the integrated circuit level, form factor issues interplay critically with packaging – and this is an often-neglected issue. Typically, integrated circuit development has focused on “chip sets” – basically, if a chip could be made to work as a chip, then the mission was accomplished. In this sort of “digital thinking,” the packaging of those chips is merely an afterthought – some plastic (or ceramic) is wrapped around the IC to protect it from the external environment. However, in RF-IC design (particularly at higher frequencies), the packaging is not an afterthought – it is **absolutely critical**, and unlike the situation with “mainstream” ICs, it must be **an integral part of the design rather than just an afterthought**.

Another key item is the development of an appropriate semiconductor technology strategy. This issue is actually much more complicated than is commonly believed. Choices must be based on providing the most design flexibility and the lowest

design risk *along with* the lowest-possible cost. All of these factors interact with the need for high data rates, simple scalability to even higher data rates, low noise, low jitter, and low power consumption; low power consumption is a factor for both battery operation **and** the reduction of heat generation, as the latter can actually perturb surrounding components. Finally, as noted above, packaging is a critical “core” issue for future RF-IC design; it cannot merely be an afterthought.

A wide variety of semiconductor technologies can be considered for very aggressive RF-IC design. This kind of evaluation is a very challenging problem, since these technologies constitute a continuum across the factors of cost, availability, and technical suitability.

Here, we will highlight a danger in this sort of evaluation. A dangerous “theology” has developed that CMOS is such a perfect technology that it can do anything and everything, and is thus always the best choice. This “theology” is dangerous because it subsumes critical thinking about the wide range of choices and trade-offs inherent in technology – and in particular how the *goals of a design* affect the choices of inputs. CMOS is one of the most remarkable technologies ever invented by humans. The primary world-beating advantages that have propelled CMOS to its (deservedly) exalted status are scaling theory [7], [8], ease of functional integration, and **low cost**; in fact, at this time the most important advantage CMOS enjoys over alternatives is its low cost.

Because of these capabilities, it is unfortunately very easy to become intellectually lazy and not want to be bothered any longer with evaluating the details of any design problem. In addition, most of the CMOS “theology” is rooted in digital design; astounding success there has given many the impression that the basic ideas (mainly related to integrating “everything” and at low cost) are thus easily applied universally. However, this “theology” runs into difficulties in analog CMOS design; these problems become extremely severe in RF design.

When one approaches RF-CMOS design, one notes that the reality is quite different from the “digital” world. The wonderful and simple principles of CMOS scaling simply do not apply in this realm. Massive integration of different functions is a digital ability which is more difficult when trying to absorb more “analog-like” functions; the situation becomes lethally scary when considering direct integration of RF functions. These disparate functions don’t co-exist well on the same silicon IC; in addition, these “small” blocks tend to have lower yield (and take more design time to get into good shape), so that a small region of the “integrated” IC destroys the yield of a larger piece of silicon.

Essentially, much of this “CMOS theology” is “digital thinking” which is trying to get into places where it

quite simply doesn't belong. This "theology" must be replaced by the use of critical and experienced engineering judgment, based on the end goals of the design under consideration.

For example, the cost comparisons – under which "CMOS is always the lowest cost way" – can be very misleading; things depend critically on what is being compared. Most comparisons of CMOS and, say, BiCMOS are based on comparing processes using the same generation of photolithography. Under this form of comparison, CMOS is obviously less expensive. However, this comparison is skewed; comparisons must be made **not** within the same lithography generation, but must be made based on the end-goals of the design. In terms of analog and RF design, bipolar transistors are easier to use in design, and are also much faster. Thus, comparisons must be made **across** process generations; particularly for RF-IC design, a *more aggressive* CMOS generation must be chosen, while a less aggressive BiCMOS process will be at least as capable. In many circumstances, a less aggressive BiCMOS process will be *less* expensive than the CMOS alternative. This is particularly true if the CMOS choice must involve use of the newest and most aggressive "bleeding-edge" process generation; "bleeding edge" CMOS technologies are very expensive, and also tend to be unstable (particularly for analog/RF design) for an extended period of time.

CMOS technology is a marvelous invention, but it is not completely perfect. Rather than lazily following popular "theology," semiconductor technology choices must be based on solid engineering judgment, with a clear understanding of the design goals which are under consideration.

V. ATTACKING THE DESIGN GAP

A final set of issues in next-generation RF-IC design must be approached with care – but it must be approached. There is a category of challenges which many are uncomfortable discussing because *these challenges are not material, but intellectual*.

A quiet catastrophe in challenging integrated-circuit design is a kind of "design gap." Basically, analog and RF-IC design methods are, at their base, "design-by-iteration" approaches; in this form, an effort in IC design is decidedly haphazard rather than systematic. Unfortunately, a good deal of the problem has come from an over-reliance (both tangibly *and* intellectually) on "tools." Modern design "tools" are a wonderful method of improving designer productivity; however, they do not (and *cannot*) improve designer understanding of fundamental aspects of circuit behavior. *If designers are not well-versed in the fundamental underlying aspects of what they are doing, this situation will not be rectified by the use of*

"design tools." Today, the industry has reached a juncture where this has actually become a serious problem – too many engineers learn how to run "tools" but know little about the realities of engineering! The end-result of this state of affairs is that engineering becomes totally tool-dependent, leaving no basis for groundbreaking innovation and creativity.

"Tooling" has the ability to turn good engineering into great engineering; however, it does **not** have the ability to turn bad engineering into good engineering. It is important to get to the level of "good engineering" first; this is a pre-requisite for the effective use of "tooling" *and* for the eventual attainment of the *great engineering* which is required for progress in many fields - including (of course) next-generation RF-IC design.

A good example of the danger of the "design gap" is the fundamental crisis afflicting CMOS technology at this time. As shown in Figure 2, the power supply voltage has been scaling with each process generation (due to reliability and power consumption concerns); however, at the same time it has not been possible to scale the threshold - due to fundamental physical limits [9] – [11]. This situation has a large number of frightening implications; however, the one we shall consider here is a simple *intellectual* problem.

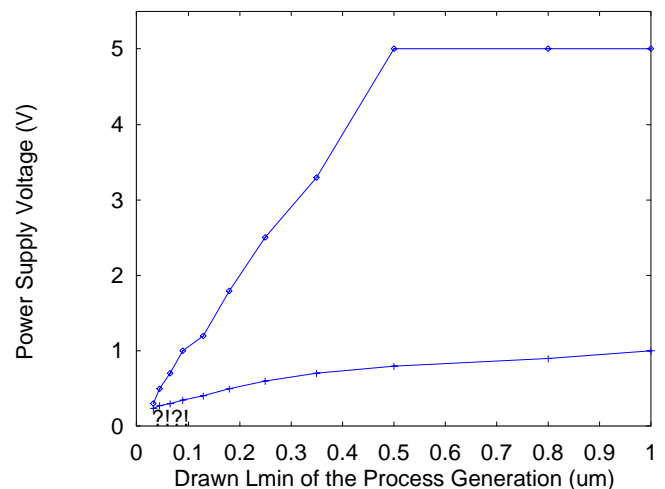


Figure 2. The power supply and threshold voltage for various CMOS technology generations.

For some forty years, all of the "rules of the road" in CMOS circuit design (particularly in analog CMOS design) have been predicated on the existence of a large gap between the power supply voltage (V_{dd}) and the threshold voltage (V_t). These "rules" were developed based on *convenience* rather than on any particular laws of physics. As Figure 2 implies, these "rules" are becoming increasingly untenable and must be replaced. Good summaries of further details may

be found elsewhere [10], [11]. However, here we will note an interesting graph of the crisis, based on the concept of the MOS “inversion coefficient” – a concept which was foresightedly elucidated by Tsvetkov [12] and Vittoz [13] more than 20 years ago. Without going into great detail here [14], it is simple to demonstrate two things: 1) The “traditional” methods of analog CMOS design set a minimum-usable value of the inversion coefficient (IC_{min}) which is independent of the technology generation; 2) It is also possible to define a maximum-available inversion coefficient (IC_{max}) for any process technology generation – and because of the scaling of V_{dd} and the non-scaling of V_t , this IC_{max} decreases with each succeeding process generation. A plot of this situation is presented in Figure 3. This graph should be heart-stopping; it basically implies that the “traditional” approaches to design are much closer to complete collapse than has otherwise been expected.

The conundrum embodied in Figure 3 is one of the most interesting ones, but it is not the only one. New methods of designing integrated circuits – *without any need to alter the process technology itself* – are badly needed. While these kinds of modernizations present obvious intellectual challenges, it also cannot be understated that there are important and tangible “business” implications. *The integrated circuit industry is, by its nature, very competitive and very capital-intensive. More advanced approaches to modern integrated circuit design have direct bearing in the marketplace. These more advanced approaches represent extremely valuable intellectual property which will be employed in ways which enhance the competitiveness and success of those companies which make them fundamental to their strategies.* These challenges have already been noticed by risk capital managers – many have noted with rue that the biggest difficulty afflicting the companies in which they have invested is the problem of “too many design iterations.”

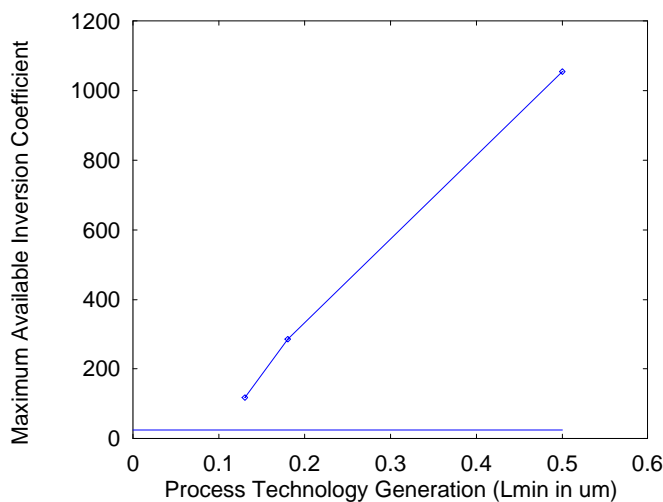


Figure 3. The maximum-available inversion coefficient in various process technology generations; the horizontal line at $IC=25$ represents the minimum usable inversion coefficient in traditional analog CMOS design methods.

VI. FROM INTELLECTUAL CHALLENGES TO SUPERIOR PRODUCTS

Based on these realities, we will note that it is possible to identify three requirements for the creation of *superior* integrated circuit products. The first is the use of a “stable” semiconductor process technology; this is a much more promising strategy than any which rely on the use of “bleeding-edge” process technology – “bleeding edge” in either specific aspects of the technology or the process generation itself. It must also be noted that there is a difference between “digital stability” and “analog stability” – processes are often suitable and “stable” for digital designs while they are unsuitable and unstable for analog designs. The second requirement is for *robust* designs – designs in which there has been a full and proper evaluation of the trade-offs, sensitivities, and risks. Too often under “iterative” design approaches, a design point is reached after a long journey and thus few questions are asked about whether this represents a “good” design point or not; in addition to the demand for further engineering time and energy as “post-processing,” the “traditional” (non-systematic) approaches simply do not provide any self-contained guidance on this matter *during the design itself*. And third, modernized design methods are required – methods which are not merely “accurate,” but which transcend the “traditional” approaches to design and thus open new capabilities *while still using the same, extant process technologies*. It is these modernized design methods which provide the **real** impetus for as seemingly-arcaic a subject as transistor models [15], [16]; viewed in this light, transistor modeling can be seen (deservedly) as a fundamental competitive advantage in the integrated circuit marketplace.

Taken together, these points embody the critical needs which must be met in a demanding, competitive environment. Design iterations must be reduced to be competitive – as design iterations waste valuable resources in time, money, and human energy. Design stability must be a critical focus item; too often, victory is declared when the first lot is a success – only to have that success mysteriously evaporate completely by the time of the tenth lot. Furthermore, modernized approaches inherently recover *simplicity* from the overwhelming and ultimately crippling growth of extraordinary complexity. One particular consequence of simplicity is so salient that it must be noted here. At this time, one of the largest single line-items driving costs in the semiconductor industry is

the cost of testing; this is a fact of life which gets very little attention. The ability to create very advanced designs *while at the same time improving the simplicity of the designs themselves* allows for major reductions in testing costs.

As a final note, there is emerging recognition that the “old ways” of doing things – of breaking the design task up into tiny, independent pieces of work which are later strung together in an assembly-line fashion – has become counterproductive and detrimental. The most notable manifestation of this realization is the increasing discussion of “design for manufacturability” (now so commonly-discussed that it carries the acronym of DFM); this is a recognition that manufacturability issues must be considered at the *start* of a design, rather than at the end as an afterthought. However, this is just a start. There is a requirement for increasing recognition that *every* aspect of integrated circuit engineering is related to every other one – there can no longer be isolated “pigeon-holes.”

Next-generation RF-IC design demands a new mindset - one which recognizes this reality. Ultimately, these issues will be sorted out via the most efficient vehicle we have – the competitive marketplace of the integrated circuit business.

VII. CONCLUSIONS

This paper has examined the present situation in high-speed wireless data communications. Despite the steadily-growing demand for wireless bandwidth, technology development in this area has become somewhat stultified – as it is focused on attempts to extend present-day technology with a variety of short-term “tricks.” The “tricks” are only made possible by paying a very serious price in other engineering aspects of the technology, such as transmission distance, power consumption, and equipment form factor. A better alternative is to take a route which uses higher carrier frequencies and demands an extensive amount of innovation; the challenges which are inherent in such an approach are manifold, demanding both tangible and intellectual innovation. Meeting these challenges will be difficult, but it is the only method by which technical simplicity and low power consumption can be reached; it is simplicity and low power consumption which will enable the development of truly portable devices, which will form the basis for true high-speed wireless peer-to-peer (WP2P) networking.

ACKNOWLEDGMENTS

The authors wish to thank all the members of the RFive Communications team for their efforts in launching this new initiative. For this paper, they particularly thank Michael Schröter and Gennady

Gildenblat.

REFERENCES

- [1] J. Borel, “General Introduction, Motivation, and Requirements for LV-LP ICs,” Proceedings of the 1993 European Solid State Device Research Conference (ESSDERC), pp. 911 - 918
- [2] D. Foty and E. Nowak, “MOSFET Technology for Low-Voltage/Low-Power Applications,” IEEE Micro, June 1994, pp. 68 – 77.
- [3] M. Levitov, “MegaFon Offers High-Speed Web Use,” *The Moscow Times*, 21 January 2005 – available on-line at: <http://www.themoscowtimes.com/stories/2005/01/21/042.html>.
- [4] “Internet Accessible Device Annual Shipments Headed Past 1 Billion,” In-Stat, January 2005.
- [5] Y. Huang, “Can Adaptive Signal Processing Do More for Modern Multiple-Access Communication Systems?,” Proceedings of the 2003 European Conference on Circuit Theory and Design (ECCTD), p. 1-15.
- [6] P. Beckmann, *A History of π* , New York: St. Martin’s Press, 1971.
- [7] C. Mead, “Fundamental Limitations in Microelectronics – I. MOS Technology,” Sol. St. Elec. vol. 15, pp. 819 – 829 (1972).
- [8] R. Dennard et al., “Design of Ion Implanted MOSFETs with Very Small Physical Dimensions,” IEEE J. Sol. St. Circ. vol. SC-9, pp. 256 – 268 (1974).
- [9] E. Nowak, “Ultimate CMOS ULSI Performance,” 1993 IEDM Tech. Dig., pp. 115 – 118.
- [10] D. Foty, “Re-Interpreting the MOS Transistor for the 21st Century: Generalized Methods and Their Extension to Nanotechnology,” Proceedings of the 21st Nordic VLSI Design Conference (NorChip), pp. 8 – 15 (2003).
- [11] D. Foty and G. Gildenblat, “CMOS Scaling Theory – Why Our ‘Theory of Everything’ Still Works, and What That Means For The Future,” Proceedings of the 12th International Symposium on Electron Devices for Microwave and Optoelectronic Applications (EDMO), pp. 27 – 38 (2004).
- [12] Y. Tsvividis, “Moderate Inversion in MOS Devices,” Sol. St. Elec. Vol. 25, pp. 1099 – 1104 (1982).
- [13] E. Vittoz, “Micropower Techniques,” in *Design of MOS VLSI Circuits for Telecommunications* (ed. By J. Franca and Y. Tsvividis), Prentice Hall, 1994.
- [14] D. Foty, “An Evaluation of Deep-Submicron CMOS Design Optimized for Operation at 77K,” Proceedings of the 12th International Conference on the Mixed Design of Integrated Circuits and Systems (MIXDES), pp. 939 – 945 (2005).
- [15] M. Schröter, “HiCUM – A Scalable Physics-Based Compact Bipolar Transistor Model,” Description of Model Version 2.1, January 2001.
- [16] G. Gildenblat et al., “SP: An Advanced Surface-Potential-Based Compact MOSFET Model,” IEEE

J. Sol. St. Circ. vol. SC-39, pp. 1394 – 1406
(2004).

Daniel Foty is presently the founder and President of Gilgamesh Associates (Fletcher, Vermont), and co-founder and Chief Technology Officer of RFive Communications (Westlake Village, California). Gilgamesh Associates is a supplier of intellectual property in the form of next generation MOS transistor models and associated “design kits,” which are used by fabless semiconductor companies for IC design in very demanding areas of analog, mixed-signal, RF, signal processing, communications, and ultra-low-power design. RFive Communications is a start-up company specializing in next-generation integrated circuit technology for high-speed wireless data communications. He also holds the position of Senior Model Development Engineer at Nanopower Technologies (Newport Beach, California), and is an Adjunct Professor of Electrical Engineering at the Georgia Institute of Technology (Atlanta, Georgia). He is also the author of the best selling book, *MOSFET Modeling with SPICE: Principles and Practice*, which was published in 1997 by Prentice-Hall and is now in its third printing. Dr. Foty has more than 15 years of engineering and management experience in the mainstream of the integrated circuit industry. He has also authored more than 70 journal articles and conference presentations, and is a frequent plenary (keynote) speaker at major international conferences throughout the world. He holds B.S. degrees in Physics and Chemistry from Bates College, the M.S. in Electrical Engineering from the University of Illinois, and the Ph.D. in Materials Science from the University of Vermont.

Klaus Runge joined the Applied Research division of Bellcore in 1988. His research interests included high-speed silicon bipolar integrated circuits, subcarrier multiplexed lightwave systems, and regenerator design for direct detection lightwave systems. In 1993 he joined the Rockwell Science Center where he was involved in the design of ultra-high speed (40+ Gb/s), as well as high-speed low-power AlGaAs/GaAs HBT circuits. At both Bellcore and Rockwell he set several world-records for integrated circuit switching speed and analog performance, in both silicon bipolar and GaAs HBT technologies. In 1995, he won the Rockwell Chairmans', and Presidents' Team Awards, for leading the design effort of a low-power lightwave SoC. In 1998 he was involved in research on RF CMOS for wireless and lightwave applications, where he set several records. In 1999 Mr. Runge helped form GTRAN Inc., where he was Director of Circuit Development (optical receiver and SiGe circuit product lines) and Technical Fellow (12.5 Gb/s SiGe, and 40 Gb/s InP) circuits. In 2001 he was Executive Director of product development for Talaria, where he was responsible for the companies InP products, and contributed to the development of the InP production technology. In 2005, he co-founded RFive Communications, where he is CEO. Mr. Runge has authored more than 57 journal/conference presentations, tutorials, invited presentations, and a book chapter on high-speed electronics for optical communications. He is a member of Eta Kappa Nu, and the IEEE. Mr. Runge holds the degrees of B.S.E.E., 1986 and M.S.E.E. (with honors), 1988 from Rutgers College of Engineering.