Driving lower fiber optical power consumption through monolithic electronic and optoelectronic integration

L. Tarof

ELPHiC, 202-2555 Michael Cowpland Dr, Ottawa, ON, Canada K2M 0M5, larry.tarof@elphic.com, 1-613-686-6770

Keywords: InP, E-O integration, monolithic integration, receiver sensitivity, power consumption, optical communication

Abstract

There has been considerable effort in both SiPh and InP on developing optical PICs for datacenter and other applications. The partitioning between optical PICs and ASICs, though, necessitates unwanted parasitics, intrinsic to the architecture itself, which limit system performance. If this boundary between optical PICs and ASICs is removed, superior performance becomes possible, which in turn can drive lower fiber optical power consumption. Some specific examples are presented

Introduction

It is well known that for datacenter, wireless and PON applications, overall power is a vital consideration. Separate from the overall carbon footprint and climate change considerations, datacenters are often located near a specific power source and have brick wall power limitations. In addition, any power consumed in transmitting, receiving or maintaining temperature locally has a multiplier effect in overall cooling of the datacenter. And received signals within datacenters or 5G often operate well in excess of 30dB from the quantum photon limit [1]. Considerable effort has gone into optical PICs, whether in SiPh or InP. SiPh has the advantage of larger wafers and greater overall infrastructure. InP has the advantage of direct bandgap native materials for light transmission and detection. Both are viable. ASICs have the maturity of a half century of Moore's Law. There are very few transistor options c.f. the plethora of potential elements in optics PICs. This means that the design flow of ASICs is quite mature whereas the design flow of optics PICs, whether SiPh or InP-based, tends to be less mature and more complex. Moreover, the ASICs which can be produced are massive in gate count – e.g. DSPs. Optics PICs, on the other hand, tend to have medium-sized element count in disparate elements, and are proceeding in gate count much slower than the Moore's Law for ASICs because of the complexity in disparate elements [2]. This segregation, though, leads often to the need to connect electrically the optical and electronic elements, and with associated parasitics at the very connection point where the signal can be the smallest or the penalty for parasitics can be the greatest. If these parasitic elements could be significantly reduced or eliminated, then performance could be improved and power within the datacenter decreased.

Discussion and Results

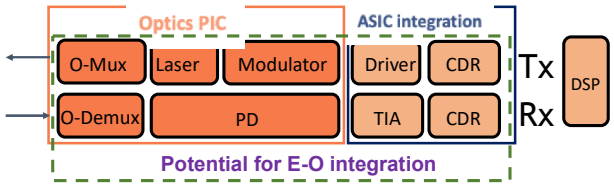

Fig. 1 Potential for E-O integration. Optics PICs and ASICs tend to have segregated functionality. E-O integration can bridge this gap, improving power consumption.

Fig. 1 shows schematically how such a fiber optical module or board partitioning looks. Those elements which are closest to the optics tend to be more analog in nature, relatively low in gate count sometimes, and more amenable to monolithic integration with optics PICs. SiPh tends to be run in more cost-effective technologies, meaning that the transistor gate length tends to be several generations behind the latest and greatest. This is because optical elements have significantly larger physical dimensions than transistors. Although it is possible to run the optics PIC functionality in low gate length Si technologies, the cost per area is usually prohibitive. In the larger gate length technologies, the gate count is necessarily limited, as is the electronic speed. And so it happens that in general both SiPh and InP-based photonics tend to be optoelectronics and optics rather than electronics. And for all its advantages, SiPh has the further disadvantage that there is no efficient direct bandgap material near 1.3 or 1.55um wavelength, meaning that light detection and most especially light generation are more difficult. It is unlikely that the high gate count of the DSP would be fruitfully reproduced in a high gate length Si technology or InP technology. However, analog functionality is quite another matter. Analog functionality is often better accomplished in bipolar than FET technology in the first place, whether in BiCMOS or InP. This makes SiGe:SiPh and InP ideal materials with which to implement analog functionality on the same substrate as the optics PIC. The present industrial state of the art for f_T is approximately 300GHz for SiGe [Zimmer] and for InP [GCS]. This enables circuits with baud rates of 25GB or even 50GB.

In the case of transmitters, the integration of a driver with electronics would save considerable power by making the transmission line more efficient. The starting point, though, may be the receiver. A typical receiver sensitivity is far from the quantum limit, as is seen in Fig. 2. In this case, PON applications are separated from non-PON applications because of the up to 32:1 split from the head end to the client end, resulting in significantly better receiver sensitivity required. It is also worth noting that 2.5G PON and 10G datacenter applications are much older than the newer 25G applications, and these newer applications are, apples for apples, generally closer to the quantum limit, although still very far away. Of further note is applications with pin PDs usually need to satisfy overload requirements of +3dBm or so incident upon the receiver, with all the system architecture considerations implied upon the pin-TIA combination.

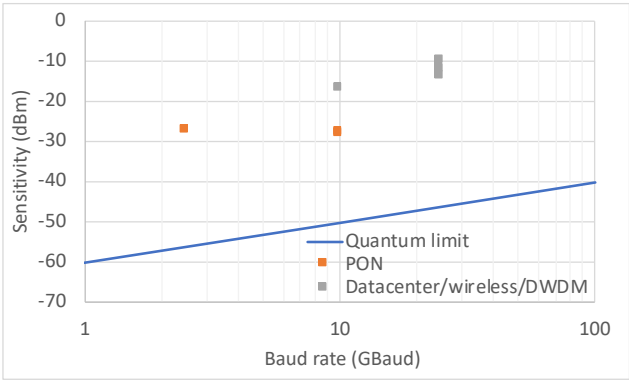

Fig 2. Sensitivity for typical applications c.f. the quantum limit.

For PON applications, an APD is typically used, which is usually 3-5dB better than a pin PD, whereas a pin PD is usually used for non-PON applications. The exception is for extended reach, where more reach than a pin PD would be desired, but because of gain-bandwidth limitation of commercially available APDs, it is challenging to use an APD at 25 GBaud or higher. In these cases, extensive DSP is often used to compensate for the poor high speed performance of APDs, but is it clear that the system limit is under strain. The usual way to compensate for this is to boost the transmitter power, but it is clear that the overall system power would be significantly reduced if the receiver could be improved at the fundamental level. All benefits to the receiver can in principle flow immediately to the transmitter launch level for a given link budget, immediately reducing the power of the overall system. The transmitter, in turn, can be made more power efficient in its own right, and building on the benefits already accrued from the superior sensitivity of the receiver.

The most common configuration for amplifying the small signal from a PD to an output level which can be useful to the outside world involves TIAs. The lower the capacitance of the pin and the TIA, the higher the transimpedance can potentially be without noise dominating. The most effective way to reduce the practical capacitance of a PD is to eliminate the bond pad capacitance. The intrinsic capacitance of the light-producing portion of the PD is necessary, but the bond

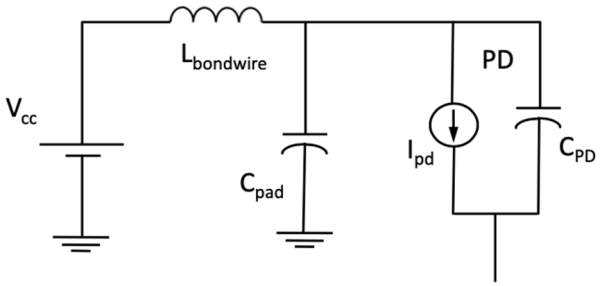

Fig. 3. Equivalent circuit for monolithically integrated PD.

pads merely serve to connect the PD to the outside world and their capacitance is merely parasitic. In this paradigm, reducing the active area of the PD, adding some complexity to the fiber alignment, is worth it, because C is no longer dominated by the bond pads. Using monolithic integration, one bond pad can be eliminated by direct interconnect between the pin PD output and the first stage TIA input. The other connection which supplies V_{cc} can be considered in the context of the bondwire, as shown in Fig. 3. In this case there is only one bond pad, connected to an external voltage V_{cc}, because the opposite end of the PD is connected monolithically to the base of the input transistor on the TIA. C_{pad} and $L_{bondwire}$ are the bond pad capacitance and wirebond inductance. The PD is represented as a current source I_{pd} in parallel with the intrinsic capacitance C_{PD}. The one bond pad is eliminated and the second one can be ignored under certain practical conditions.

The most important consideration for the second bond pad is the resonance frequency relative to the bondwire, which is calculated in Fig. 4 as a function of frequency for a typical

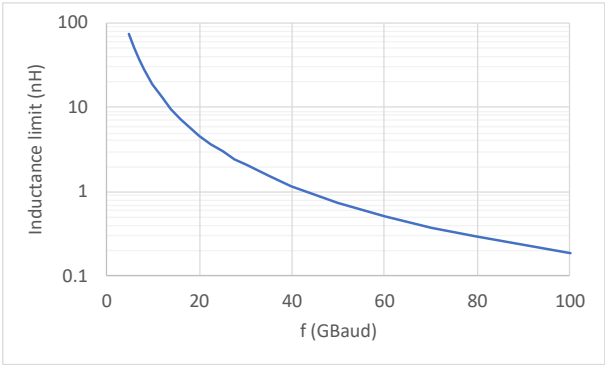

Fig. 4 Wirebond inductance limit as a function of resonance frequency for circuit in Fig 3. The assumed bondpad capacitance was 15fF.

bond pad capacitance of 15fF, which capacitance is a typical value based on vendor data. Below this limit the electrical effect of the bond pad can be ignored. For typical wirebond

lengths and diameters, in practice this means that no attention needs to be given to this below approximately 50GBaud, which is helpful for today's baud rates.

For baud rates for ~50GBaud and below, therefore, it is practical to consider monolithic integration of the PD and TIA. A further consideration, for long-reach and for PON applications, where APDs are already needed, is that a pin PD will, all else equal, always be faster than an APD because in the pin PD there is only one transit time for carriers, usually only part of the absorption region, whereas in the APD the initial carriers must travel from the absorption region to the multiplication region, be multiplied, and then the secondary carriers must return through the full distance through the absorption region to reach the contact [Campbell??]. If the pin-TIA connection can be made quieter, then there may be need to "force" an APD into the system architecture for sensitivity considerations, and also suffering the associated bandwidth and therefore DSP power penalties.

In summary, the foregoing demonstrates that there are parasitics benefits to monolithic integration of the PD and TIA, but do not yet demonstrate that there is an overall receiver benefit. Perhaps the most clear-cut way to evaluate this is by considering a single PD and first-stage TIA and evaluating the SNR. The associated SNR is the single largest indicator of eventual sensitivity at the higher voltage end of the circuit.

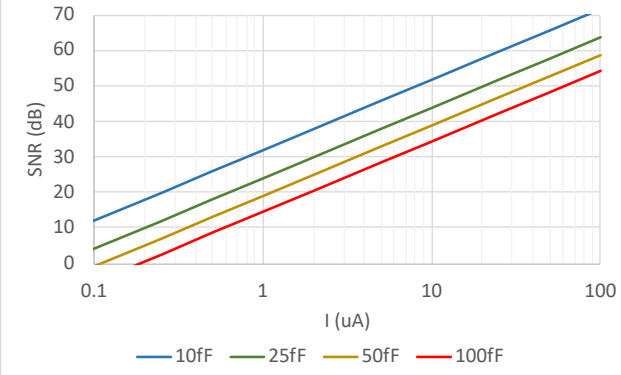

Fig. 5. SNR (dB) vs input photocurrent I for different C values.

Fig. 5 shows the SNR vs input photocurrent for different capacitance C values. In this case, an example transimpedance of 40dB-ohms was used. This could be made even higher because of the lower Johnson noise associated with the lower input capacitance, which is a key benefit of the monolithic integration approach, but even 40dB-ohms will serve to demonstrate the point. Red is "hotter" (more C) moving through the spectrum to blue, which is "colder" (less C). In this case the photocurrent is average value and an infinite ER was used for simplicity, meaning the zeros have no current. It is clear that the SNR increases by 20dB per decade of input photocurrent, which makes intuitive sense. As a design curve, Fig. 5 can be used for any given SNR, indicative of an incumbent SNR solution, drawing a horizontal line to see how much improvement in the photocurrent, and therefore input power, can be had by lowering the PD capacitance by means of monolithic integration. As a reference, for a typical quantum efficiency of 75%, 1uA average photocurrent at 1.3um is equivalent to -29dBm of input power, so this power scale goes from approximately -39..-9dBm of input power spanning three decades. As a ballpark figure, roughly 10dB of SNR should translate to sufficient sensitivity at a BER of 1E-12.

It is evident from Fig. 5 that although the C spacing between curves increases from blue towards red, the curves actually get closer together. This effect is shown more explicitly in Fig 6, which shows SNR vs C for different values of average input photocurrent. For PDs with high C there is limited benefit to using this technique, but for small geometry PDs for which the bond pads form a significant fraction of the capacitance, as is the case for 16um and 20um active area PDs today [4] and even smaller which can accommodate an optical fiber for packaging, there can be significant benefit. In this case, all else equal, an improvement in SNR of approximately 8dB can be expected for the same input power by lowering C or, equivalently, an improvement of approximately 4dB in input sensitivity for the same SNR from a practically achievable capacitance reduction. This improvement can be stacked with other improvements to the PD itself which might be realized because of the ability to monolithically integrate, and also with the accrued benefits of no external contact from the PD to the TIA, which is in general a source of noise for many reasons, including acting as an antenna for picking up other nearby signals as crosstalk as an integral part of the packaging considerations.

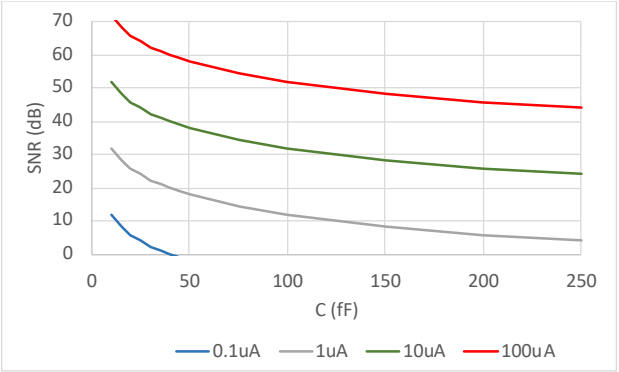

Fig 6. SNR (dB) vs C for different input photocurrents values.

It is worth considering why this work is not done commercially in general. It is clear that the idea of integrating a pin PD and TIA has been around for a long time [6]. It is also worth noting that in the past it is not clear that there was a performance advantage from monolithic integration. There is generally a cost and often a density advantage from monolithic integration, but monolithic integration is often a compromise among the integrated elements. Not so in this case, where clear performance advantages are observable.

ELPHiC has carried out more detailed modeling for a full pin-TIA circuit, and considering all the packaging parasitics. In the package, modeling shows that for a 10G PON application, significantly better than -30dBm could reasonably be expected in the TO can, meaningfully outperforming a typical APD-TIA incumbent solution, even after taking into account packaging considerations. This is particularly important in PON applications because the next gen PON will either be 25G or 50G in baud rate. For 25GBaud, an APD is conceivable if difficult to implement. For 50GBaud, an APD is out of the question. For this future application, if a 32:1 split ratio is to be preserved, there must be either an increase in transmitter power, which is highly undesirable, or an improvement in receiver technology architecture, which monolithic pin-TIA can offer. In this case one can avoid a significant increase in transmitter power and all the associated overhead despite the significant increase in baud rate.

Similarly, for fronthaul 5G applications ELPHiC modeling shows that a pin-TIA monolithically integrated solution is expected to outperform incumbent solutions. A receiver sensitivity better than -20dBm could reasonably be expected in the TO can, even after taking into account packaging considerations.

Note that both the 10G and 25G examples reported here are a significant improvement over incumbent solutions as shown in Fig. 2.

APPLICATIONS

For APD-based applications today, such as PON or some of the extended reach 25G applications, it is conceptually easy to use a pin-TIA monolithically integrated combination as a substitute for APDs. A further system architecture advantage is the reduced power supply need, as an APD requires tens of volts whereas the pin-TIA combination requires only the standard 3.3V. This can greatly simplify board design.

For pin PD-based applications, monolithic integration can improve the sensitivity, which in many cases is an advantage either, from a system link budget perspective or to permit lower transmitter power. If the overload requirement is to be preserved, this implies an increased dynamic range requirement on the TIA. An important consideration is whether is it more important to preserve the overload requirement or to maintain the TIA dynamic range, which would imply a lower overload limit.

MSA standards already exist for many 10Gb/s and 25Gb/s applications. However, MSA standards have yet to be finalized for many 25GBaud and 50GBaud applications. A monolithically integrated receiver can help drive a lower transmitter spec, saving fiber optical power going forward.

CONCLUSIONS

In conclusion, it is possible to repartition the overall fiber optical module or board chipset architecture in such a way as to permit the monolithic integration of optical PICs and electronics for the more analog electronics or where the gate count is not excessively high. In that case, the overall power in a fiber optical system can be substantially reduced first through improved receiver sensitivity, which permits lower transmitter power, and then through improving the transmitter itself through analogous means. Some specific example simulations have been shown which demonstrate this.

ACKNOWLEDGEMENTS

This work is really a collaboration among many, to whom the author is indebted. In particular, medeling and concepts from Professor Rony Amaya (Carleton University), Jim Hjartarson, Andre Hagley, Rick Clayton, Shakeeb Abdullah, Wenyu Zhao and Rene Nyangezi have been most illuminating.

REFERENCES

[1] L. Tarof *Fundamental and Technology Ecosystem Considerations for Photonics ICs and Datacenter Architectures*, 2019 Optinet, Paper T72, June, 2019.

[2] L. Tarof, *How advanced packaging can help align the evolution of photonic ICs with Moore's Law*, Laser Focus World, pp.20-21, Mar, 2017

[3] Thomas Zimmer et al, *SiGe HBTs and BiCMOS Technology for Present and Future Millimetere-Wave Systems*, IEEE Journal of Microwaves 1 (1), pp.288 – 298 (2021). 10.1109/jmw.2020.3031831 hal-03111157

[4] As one example of what is commercially available: *http://www.gcsincorp.com/dedicated_pure-play wafer foundry/InP%20HBT%20Technologies.php*

[5] J. C. Campbell, "*Recent advances in avalanche photodiodes*, IEEE Journal of Lightwave Technology **34**, (2), pp. 278-285 (2016), doi: 10.1109/JLT.2015.2453092.

[6] S.Chandresekhar et al., *Eight-Channel p-i-n/HBT Monolithic Receiver Array at 2.5 Gb/s Per Channel for WDM Applications*, IEEE Photonics Technology Letters **6** (10), pp.1216-1218 (1994).

ACRONYMS

APD: Avalanche Photodiode
ASIC: Application Specific Integrated Circuit
CDR: Clock and Data Recovery
DSP: Digital Signal Processor
ER: Extinction Ratio
HBT: Heterojunction Bipolar Transistor
O-Demux; Optical Demultiplexer
O_Mux: Optical Multiplexer
PD: Photodiode
PIC: Photonics Integrated Circuit
SiPh: Silicon Photonics
SNR: Signal to Noise Ratio