FEATURED IN THIS ISSUE:

GSA SITS DOWN WITH BEHROOZ ABDI, CEO, INVENSENSE
RAISING THE LEVEL OF ABSTRACTION FOR HETEROGENEOUS PROGRAMMABLE DEVICES
ISSI MEMORY SOLUTION – FOR 400G+ LINE RATE
TECHNOLOGY SELECTION IMPLICATIONS INTENSIFY AND OPTIONS ARE LIMITED
Performance. To get it right, you need a foundry with an Open Innovation Platform® and process technologies that provides the flexibility to expertly choreograph your success. To get it right, you need TSMC.

It is TSMC’s mission to be the Trusted Technology and Capacity Provider of the global logic IC industry for years to come. In this regard, TSMC assures your products achieve maximum value and performance whether your designs are built on mainstream or highly advanced processes.

Product Differentiation. To drive product value, you need a foundry partner who keeps your products at their innovative best. TSMC’s robust platform allows you to increase functionality, maximize system performance and differentiate your products.

Faster Time-to-Market. Early market entry means more product revenue. TSMC’s DFM-driven design initiatives, libraries and IP programs, together with leading EDA suppliers and manufacturing data-driven PDKs, get you to market in a fraction of the time it takes your competition.

Investment Optimization. Every design is an investment. Function integration and die size reduction help drive your margins; it’s simple, but not easy. TSMC continuously improves its process technologies to get your designs produced right the first time.

Find out how TSMC can drive your most important innovations with a powerful platform to create amazing performance. Visit www.tsmc.com
ACCELERATE THE GROWTH AND INCREASE THE RETURN ON INVESTED CAPITAL OF THE GLOBAL SEMICONDUCTOR INDUSTRY BY FOSTERING A MORE EFFECTIVE ECOSYSTEM THROUGH COLLABORATION, INTEGRATION AND INNOVATION.

- Address the challenges and enable industry-wide solutions within the supply chain, including intellectual property (IP), electronic design automation (EDA)/design, wafer manufacturing, test and packaging
- Provide a platform for meaningful global collaboration
- Identify and articulate market opportunities
- Encourage and support entrepreneurship
- Provide members with comprehensive and unique market intelligence

MISSION AND VISION STATEMENT

CONTENTS

3 Potential Futures: The Drivers of Change & the Emergence of Alternative Manufacturing Models in the Semiconductor Industry
Stephen Rothrock, President & Managing Principal, ATREG, Inc.

8 Raising the Level of Abstraction for Heterogeneous Programmable Devices
Desh Singh, Director of Software Engineering, Altera

12 Technology Selection Implications Intensify and Options are Limited
Christophe Maleville, Senior Vice President, Digital Electronics Division, Soitec

14 ISSI Memory Solution – for 400G+ line rate
Anandarup Bagchi, Director, Strategic Marketing, ISSI

17 Monolithic 3D IC: an Emerging Technology to Keep Us on Moore’s Law
Zvi Or-Bach, President & CEO, Monolithic 3D, Inc.

20 Conflict Minerals in the Semiconductor Industry
Ron Jones, CEO, N-Able Group

23 Innovative Solutions to Data Center Issues
Ariel Ueng, Market Associate, Angel Orrantia, Business Development Director, SKTA Innopartners

ARTICLES

IN EVERY ISSUE

6 Private Showing

11 Industry Reflections

INTERESTED IN CONTRIBUTING TO THE GSA FORUM?

To contribute an article, contact: Kate Olsen, GSA Forum Executive Editor, kolsen@gsaglobal.org. To advertise, contact: Nicole Bowman, Senior Marketing Manager, nbowman@gsaglobal.org. • GSA • 12400 Coit Road, Suite 650, Dallas, TX 75251 • phone 888.322.5195 • fax 972.239.2292

Copyright 2013 Taiwan Semiconductor Manufacturing Company Ltd. All rights reserved. Open Innovation Platform ® is a trademark of TSMC.
Accelerating Innovation

It’s our commitment to our customers and the core of Synopsys’ DNA. As the complexity of electronics grows exponentially, engineers worldwide look to Synopsys’ leading software, IP and services to help them design the chips and systems that are increasingly impacting everyone, everything, everywhere.

Learn more at www.synopsys.com/ai

What Will You Design Next?
The semiconductor industry is finding itself in uncharted waters as it continues to morph and mature from the exciting, fast-growing segment it was in its early stages. While change can be a painful process for an industry, it is also an opportunity that companies can embrace to position themselves ahead of the curve and thrive. This article examines the macro forces transforming the industry and explores the rise of alternative manufacturing models.

MACRO THEMES IMPACTING THE INDUSTRY

Industry slowdown, with the mobile market driving the majority of growth. Total revenue for the semiconductor industry peaked at $310 billion in 2011. From 1985 to 2005, revenue grew at a compound annual growth rate (CAGR) of 12.8 percent. However, between 2005 and 2012, that growth rate dropped to just 3.4 percent and current projections show no substantial improvement on the horizon as most industry segments experience anemic growth. The exception to the rule is the $10 billion mobile market, which is projected to exhibit more than 20 percent CAGR between 2011 and 2016.

Rising costs of leading-edge. The capital requirements for creating new processes and building new fabs have gone up exponentially over the last decade. In 2001, the total costs for leading-edge process and fab development (0.13μm) amounted to between $1 billion and $2 billion. These costs have risen to approximately $7 billion for a state-of-the-art facility (22nm).

Instability of the foundry model. The foundry market is experiencing its largest-ever bifurcation. TSMC is becoming immensely profitable by dominating all advanced technology production while the rest struggle to achieve sufficient loading to generate consistent positive profits. In 2012, TSMC’s net income was more than 17 times greater than the cumulative net income of the rest of the foundry sector.

POSSIBLE IMPLICATIONS FOR THE INDUSTRY

A less attractive fabless operating model. The fabless segment has been the darling of the semiconductor industry as a result of its rapid growth and reduced capital requirements compared to traditional integrated device manufacturers (IDMs). While fabless companies do not have the opportunity to benefit from the virtuous cycle of engineering and design intrinsic to an IDM and, in certain instances, have a time-to-market disadvantage, several firms have achieved enormous success by relying on dedicated foundry partners for manufacturing. However, the fabless-foundry dynamic will shift greatly in the near future as fabless companies attempt to prevent over-reliance on an increasingly limited number of suitable manufacturing partners.

Potential capacity constraints, reduced negotiating leverage, as well as geographic and natural disaster risks begin to pose real threats to the livelihood of companies outsourcing most or all manufacturing to the few foundry players that remain competitive. This new paradigm has not gone unnoticed by Wall Street – the continued compression of valuation multiples among industry segments, as illustrated in Figure 1, provides unambiguous evidence of investors no longer attributing a premium to fabless companies.

Figure 1: Compression of Enterprise Value to Revenue (EV / Rev) Multiples Among Industry Segments

Source: Company Financials
From a shareholder’s perspective, TSMC has outperformed the industry, generating significantly higher total shareholder return (TSR) over the last six years. Figure 2 shows TSMC’s TSR relative to the largest fabless companies as well as Intel, the bellwether IDM.

*Note: TSR is calculated as equity appreciation with reinvested dividends.

As the growing market dominance of a single foundry continues to add downward pressure on margins for even the largest fabless players – particularly at advanced process nodes – companies would do well to reexamine their dependence on outsourced production.

The rise of the branded OEMs. Large consumer product original equipment manufacturers (OEMs), such as Amazon, Apple, Google, Microsoft, and Samsung, are seizing the dominant share of the mobile market – the only sizable, fast-growing semiconductor segment. This allows them to capture supernormal profits and amass massive cash resources, often at the expense of semiconductor suppliers, as shown in Figure 3.

Due to their heavy dependence on the electronics supply chain, OEMs are taking steps to vertically re-integrate to increase their control over it. Utilizing colossal cash reserves and market power, OEMs will strive to ensure the component supply they depend on through any means necessary, including acquiring suppliers. Given these factors, as well as the unrelenting pressure on all participants in the semiconductor supply chain, the trend of vertical re-integration is likely to continue in the near future.

**POTENTIAL FUTURE SCENARIOS**

To consider what the future might look like, we have chosen to look at the relative dominance of TSMC and the extent to which the major OEMs further re-integrate and how one or two companies dominate the market for smart devices. Looking at these variables along two primary axes allows for the examination of four potential futures for the semiconductor industry, as illustrated in Figure 4.

---

See Potential Futures page 27
We also provide:

- A match with a **Strategic Partner**, an established public company, for collaboration and acquisition of latest intelligence and related intellectual properties.
- Access to **VCs** – SK’s internal VC and other affiliated top tier VCs to secure Series A funding.
- **Business plan development** – Innopartners’ professionals work together with entrepreneurs to develop a solid business plan to appeal to both strategic sponsors and VCs.
- Dedicated **office space** with secure access and shared conference rooms, lab, media room.
- **Professional services** including legal counseling (startup formation, patent filing, contract review), accounting and finance, marketing, recruiting, administration and IT (video conferencing, broadband access, mail server, storage server, development tools such as EDA).

While most venture capitalists are running away from semiconductors, we are actively recruiting entrepreneurs with great ideas in semiconductor and enterprise networking.

Innopartners is focused on bringing the most innovative and inspiring concepts in nanotechnology to life. Examples of subject matter eligible for funding by Innopartners include:

- Semiconductor / Components
- Device Solutions
- Telecommunication/Network
- IT
- Healthcare Device / BioInformatics
- Emerging Technologies

If you are an entrepreneur with great ideas in semiconductors or networking, **JOIN US**! We are revolutionizing the way hardware companies get funding. You bring your idea, we do the rest.

**We provide initial funding up to $1 Million**

**Interested? Find us at**

[www.SKTAinnopartners.com](http://www.SKTAinnopartners.com)
ranging from fully managed teams residing at one
mode. Our engagement model is very flexible,
augment our client’s existing staff in an on-site
around the world, or we can provide a team to
products, either through our design centers
and implementation of their next generation
experience in product design and services.
seasoned management team, with many years of
management, cloud computing, security and
computing, wireless/mobile, multimedia, power
market sectors, such as storage, networking/
and Software.
well over 700 design engineers, delivering design
solutions in SoC Design, Analog/Mixed Signal
and Software.
With the broad set of skills possessed by our
experienced engineering team, we service many
market sectors, such as storage, networking/
computing, wireless/mobile, multimedia, power
management, cloud computing, security and
Internet of Things.
Our experienced engineers are backed by a
seasoned management team, with many years of
experience in product design and services.
We assist our customers in the design
and implementation of their next generation
products, either through our design centers
around the world, or we can provide a team to
augment our client’s existing staff in an on-site
mode. Our engagement model is very flexible,
ranging from fully managed teams residing at one
in the 3D model to quickly assess yield and
reliability risks. More ground-breaking advanced
features are coming in the next major release of
SEMulator3D in early 2014. Combining these
new features with Virtual Metrology and the
Expeditor batch modeling tool, SEMulator3D
provides a full virtual fabrication platform for all
stages of technology development.
In addition to its focus on process
development, MEMS design is Coventor’s area
of traditional strength. Its CoventorWare® and
MEMS+® products are used by the world’s
leading developers of MEMS devices to design
with unmatched accuracy, performance and
capacity. IN October it released version 4.0 of
its MEMS+ suite, which enables the co-design
of MEMS and ICs in an advanced modeling and
simulation environment that connects with
popular IC design tools.

“Coventor values its role within the GSA because it
connects the company with the key markets it serves,
namely fabless IC companies and IC
manufacturers -
both IDMs and foundries. The access to information,
input channels for important technical and business
issues, and general networking opportunities available
through GSA are important for staying in touch
with customers and partners. Coventor addresses key
requirements of the fables and foundry markets.
For fabless, its tools provide detailed insight into
manufacturing process to reduce traditional build and
test cycles. For foundries and other manufacturers (both
MEMS and IC), it’s tools enable a much more predictive
and accurate way to develop new process technology.”

— Michael J. Jamiołkowski, CEO of Coventor
Michael J. Jamiołkowski – CEO
Dr. David M. Fried – CTO
Dr. Stephen R. Breit – VP, Engineering
Tom Flynn – VP, Sales and Business Development
1000 Centregreen Way, Suite 200
Cary, NC, 27513
(T) 919-854-7500
(W) www.coventor.com

Synapse Design delivers Engineering Excellence
to the largest semiconductor companies around
the globe. We are headquartered in Silicon Valley,
but truly have a global presence, with design
centers not only in North America, but also in
Europe, China and India. In total, we number
well over 700 design engineers, delivering design
solutions in SoC Design, Analog/Mixed Signal
and Software.
With the broad set of skills possessed by our
experienced engineering team, we service many
market sectors, such as storage, networking/
computing, wireless/mobile, multimedia, power
management, cloud computing, security and
Internet of Things.
Our experienced engineers are backed by a
seasoned management team, with many years of
experience in product design and services.
We assist our customers in the design
and implementation of their next generation
products, either through our design centers
around the world, or we can provide a team to
augment our client’s existing staff in an on-site
mode. Our engagement model is very flexible,
ranging from fully managed teams residing at one
of our design centers or customer’s sites to our
engineers located on-site and managed solely by
our customer’s management team.
We provide low cost off-shore teams where
that is an appropriate model, or partial on-site/
off-shore solutions if that is a better option.
At the end of the day, it is all about helping to
create successful products for our customer, and
we apply a “Will Do” attitude towards making
that happen.

“We joined GSA because of the access it gives us to
the major fab, library, IP, EDA and chip design
companies. Through GSA we have the opportunity
to network with the top management at these
companies, and to initiate a discussion with them
about ways in which Synapse Design can assist in
the design of their most challenging products.”

— James Hemeon, VP, Worldwide Strategic Accounts
Satish Bagalkotkar – President/
CEO & Co-Founder
Devesh Gautam – COO & Co-Founder
Charlie Kahle – CTO
James Hemeon – VP, Worldwide
Strategic Accounts
Tom King – VP, Business Development
Hem Hingarh – VP, Engineering
150 S. Almaden Blvd, Suite 1380
San Jose, CA 95113
(T) 408-850-3640
(W) www.synapse-da.com
The great minds of the past were driven by their passion to innovate. Today, the drive to innovate continues with a different set of time, budget and resource constraints. With Mentor Graphics innovative design solutions, teams are free to develop and optimize creative ideas effectively and more efficiently. Release your imagination and realize your greatest potential with Mentor.

To learn more visit us at www.mentor.com.
The Level of Abstraction for Heterogeneous Programmable Devices

Desh Singh, Director of Software Engineering, Altera

In the past 20 years, FPGA technology has evolved at an incredibly rapid rate. In the early ‘90s, the primary use of FPGAs was for a relatively small amount of glue logic that served to integrate discrete system components. Today, FPGA devices possess massive amounts of programmable logic capacity and also integrate custom ASIC and hardened processor technology. These components essentially form a heterogeneous SOC that is suitable for the implementation of entire applications.

One of the key challenges faced by the application developer is to manage the implementation complexity of mapping algorithms to the FPGA’s heterogeneous fabric. Traditionally, the programmable logic elements in the FPGA fabric are exploited through the use of HDL like VHDL or Verilog. Processor systems are programmed in software languages and often require interaction with operating systems and device drivers. Integrated ASIC technology is used for commonly used fixed function components such as PCI-express and external memory interfaces. These components typically interact with both hardware and software. It is evident that traditional methodologies require different design styles to efficiently use each of the FPGAs heterogeneous components. In this article, we will look at a unified design environment for FPGAs based on an industry standard language (OpenCL) for heterogeneous platforms. The use of OpenCL allows the developer to describe their applications in a single language and partition subtasks in such a way to efficiently exploit the different portions of the FPGA fabric.

The biggest challenge of designing for FPGAs is that the programmable logic fabric of an FPGA processes information in a manner that is quite different from traditional processor based systems. A microprocessor contains a fixed data-path consisting of functional units (FU) and register files that are connected with a fixed topology. Software programs are expressed using sequences of instructions that serve to produce control signals that coordinate data movement and transformation operations through the functional units and register files. In contrast, the programmable fabric of the FPGA consists of a blank slate of programmable logic elements (LE) surrounded by a configurable routing network. These LEs along with the configurable routing network can be used to implement hardware circuits of arbitrary complexity. The ability to implement any hardware circuit has profound effect on the way that we can execute a software program using the programmable logic fabric.

To illustrate this concept, let us look at a concrete example. Suppose that we are given a very simple microprocessor architecture shown in Figure 1. This basic processor has a program counter (PC) that feeds a Fetch unit which is responsible for grabbing the instruction at a particular PC address. This Instruction will typically contain an Op-code along with other data that defines the action to be taken by the microprocessor. The actions depend on the functional units contained in the microprocessor. In our example, the microprocessor has an ALU that is capable of implementing a multiply or addition. It also has dedicated units for fetching data values from memory addresses (Load unit) and for storing data values to memory addresses (Store unit). Now consider the following single line program:

Mem[100] += 42 * Mem[101]

This program simply increments the contents of memory location 100, by 42 times the contents of memory location 101. This simple program can be expressed by a sequence of high level assembly language instructions:

R0 ← Load Mem[100]
R1 ← Load Mem[101]
R2 ← Load #42
R2 ← Mul R1, R2
R0 ← Add R2, R0
Store R0 ← Mem[100]
First the register $R0$ stores the contents of memory location 100. Similarly, $R1$ stores the contents of memory location 101. $R2$ is assigned to an immediate value of 42. With all of the input data and constants loaded into registers, we can proceed to multiply $R2$ by $R1$ and add that value to $R0$. Finally, the resulting value can be written into memory location 100. To execute these instructions on a processor, the data-path shown in Figure 1 is reused by each of these six instructions in six consecutive time steps.

Figure 2: Replicating Data-paths and Optimizing

![Diagram of data-path replication and optimization](image)

Suppose that instead of reusing the same data-path in a time multiplexed fashion, we simply replicate the data-path for each instruction. This replication is shown in Figure 2(a) below. Each instruction has a dedicated data-path responsible for executing just one instruction. The orange coloring indicates the portions of the data-path that are needed for each instruction. Since the data-paths only have one particular function they need to implement, we can now significantly optimize each data-path circuit. For example, there is no longer a need to fetch instructions from memory as each data-path can be assigned to its own dedicated instruction. For data-paths that are performing operations such as loads or stores there is really no need to have multiple or addition functional units. These arithmetic units, along with the fetch unit, can simply be stripped away as shown in Figure 2(b). Similarly, for data-paths dedicated to arithmetic operations there is no need for load or store functionality. These can be stripped away as shown in Figure 2(c). Finally, these independent data-paths must be wired together to communicate information in order to execute the original software program. Register values are simply propagated from data-path to data-path to pass on partial results as shown in Figure 2(d).

The communicating data-paths shown in Figure 2(d) represent a hardware circuit that can be used to implement the functionality of the original software program. This circuit can be further optimized to strip away all unused register values. Furthermore, there is no requirement that the resulting circuit must have the same latency as the original software implementation on a processor. Thus, we are free to reschedule and ret ime this circuit into a more efficient data-path representation shown in Figure 3. This represents a circuit which is dedicated to implementing the functionality of the original software program. It is highly optimized for this particular function and only this function. There are no wasted functional units, registers or unnecessary connectivity. The hardware is completely customized exactly to the software program being implemented.

Figure 3: Fully Customized Data-path

![Diagram of fully customized data-path](image)

The power of the programmable FPGA fabric is that it can implement arbitrarily complex circuits such as data-paths that are customized for a given application. As stated earlier, traditional design flows for implementing circuits on programmable fabric is done through hardware description languages. These languages are used to describe digital circuits. One must describe the computation as a composition of Boolean functions and registers where the designer must think explicitly about the subset of operations that occur in each clock cycle. This type of detailed circuit design is tedious, error prone and difficult to scale as devices grow larger. In addition to the design of the data-path hardware itself, this circuit must also be interfaced to the external world so that data can be processed and the results communicated back to a larger system. For example, off chip memory buffers are often stored in SDRAM which require vendor specific IP cores that implement the memory controllers and low level physical interfaces. While these cores are provided by vendors and third parties, interfacing to these cores require detailed knowledge of memory characteristics, data access patterns, buffering and arbitration in order to create an efficient interface. Thus, the design time for creating a custom hardware data-path and its associated interface circuitry can grow quite expensive using the traditional HDL techniques.

Given the complexities of traditional design description methodologies, Altera began to investigate alternative design description strategies for designers who wish to use a higher level of abstraction beyond traditional HDL techniques. Instead of forcing a designer to manually create HDL for their custom data-paths, it is evident that a superior solution would be to automatically create customized data-paths from a software program using the techniques shown in Figures 2 and 3. However, the selection of the software language for design description is very important as the language has to be powerful enough to express the functionality of today’s increasingly complex applications.
Microsemi has been providing leading-edge semiconductor system solutions differentiated by power, security and reliability to high-value, high barrier-to-entry markets for more than 50 years. And we’re just getting started.

We’re committed to making the investments required to build on our leadership positions in the markets we serve: Communications, Defense & Security, Aerospace and Industrial.

Learn more at www.microsemi.com
In this interview, GSA sits down with Behrooz Abdi, CEO of InvenSense, to discuss his outlook on the expanding MEMS industry and InvenSense plans to capitalize on its growth.

**Q:** InvenSense’s product portfolio is vastly different from your previous stops which include QCOM, RMI, and NetLogic, yet InvenSense was named #1 Semi by Forbes in your first year as CEO. How have your previous successes enabled you to guide InvenSense to such rapid success?

**A:** I was fortunate in the fact that InvenSense was already a strong company on the right trajectory with a lot of talent. All I had to do was capitalize on that momentum and keep us focused on what we do best - creating great systems based on great technology, great silicon, and great software that solves an end to end problem for the customer.

**Q:** What has been the biggest advantage of your digital background that you’ve been able to leverage at InvenSense?

**A:** I think there are two things. One, I already had great market exposure and understanding of InvenSense’s core markets (CE, automotive and industrial), so the learning curve was not steep. Two, I came in with a firm understanding of product cycle development, design, and market timing. By bringing these two things to the table, I can really focus on aligning our abilities to meet our customer needs.

**Q:** InvenSense recently acquired the MEMS microphone business from Analog Devices. How do you see this acquisition strengthening your product portfolio?

**A:** Long term, InvenSense is focusing on enabling “AlwaysOn” context aware devices - devices that learn customer’s habits and become more predictive around the customer’s needs. This ultimately helps the customer be more productive by analyzing the customer’s location, what they are doing, and their daily routine. Our goal is to combine our motion and audio technologies in order to create a more seamless and intuitive interaction between customer and device.

Mid to short term, the acquisition gives our company greater context to our customers. If you look at the industry, microphone content in devices is increasing and we believe we have the technology and experience to contribute to this growing product line. Prior to the acquisition, we had already been working on our own microphone development. However, we felt that the expertise the team from ADI possessed would help accelerate this development and get us to market sooner.

**Q:** In your 10 years of existence, InvenSense has continually decreased the sensor size while increasing sensor integration. Where do you see InvenSense’s next step of innovation?

**A:** If you look at InvenSense’s chips they have as many as 9-axis’ of motion sensing, AMS and front-end digitalization, DSP software, etc. - very complex devices that must detect very small signals and that must calibrate time after time. As one continues to push the complexity of these devices, offsets, noise performance, and minute signal detection are issues that have to be addressed and as the number of interacting MEMS devices increase, the calibration and fusion become exponentially more difficult. All this to say, InvenSense will continue to drive innovation through integration - enabling seamless operation between our MEMS devices and our customers systems through improved digital circuitry and software is what we’re about. We can’t just focus on the integration between sensors; we have to focus on how they interact on a larger scale.

**Q:** A key enabler for design houses is a quick and affordable engineering sample. In the semiconductor world, shuttle wafers allow smaller companies to receive advanced technology samples at a viable price point. NF Shuttle is InvenSense’s MEMS version of a CMOS shuttle. What are the benefits and successes of the NF-Shuttle program?

**A:** This program is very beneficial to InvenSense in terms of next generation technology development. While we do our own experiments on a shuttle, we also open it up to researchers, smaller companies, and universities. This allows the customer to innovate on InvenSense’s platform and allows us a broader scope to see what new technology trends may be occurring. It allows InvenSense a view of the future by providing a technology platform to the market. This also facilitates a way for InvenSense to participate in the customer’s technology through different avenues such as partnerships or investments.

**Q:** At GSA, we talk a lot about the ability of MEMS products to give new life to 6” & 8” CMOS wafer lines. What are your thoughts on this and do you see this as an opportunity for your foundry partners?

**A:** Yes, of course. A lot of InvenSense’s technology runs on .35
In today’s semiconductor industry the trend of technology innovation is clearly being driven by mobile consumer applications and the requirements for ever increasing richness in the user experience. The consumer driven market direction places many challenges on our industry in terms of cost, time to market, power and performance. Also, consumer based markets are inherently fast moving in terms of market turns for new functions and features. The implications to the fabless and foundries of hitting or missing a market window are more significant than ever, where small moves in market share gain or loss can significantly make or break a mid tier technology supplier and significantly influence the market cap of any publicly traded company.

Today most of the fabless and foundries are ramping their initial offerings at the 28nm technology node in either LP or HP technology variants. As the market moves toward the end of Bulk CMOS at 28/20nm; due to the short channel effects and increased leakage induced parametric yield loss and escalating costs with minimal scaling advantages, the industry collective is forced to no longer rely on traditional technology scaling to deliver increasing value to the consumers.

The challenge is that beyond 28nm Bulk CMOS conventional thinking provides us three unattractive options; do nothing and stay at 28nm Bulk and innovation stagnates, move to 20nm Bulk CMOS which does not deliver any clear advantages in terms of power/performance/cost and move directly to 16/14nm FinFET and take on significant risk in terms of design execution and cost unknowns.

Assessing these three options one can quickly dismiss the first option which is, do nothing. This choice is a clear market exit strategy since the semiconductor industry is a never ending, fast moving, innovation based industry and halting will quickly result in a company’s failure.

As we look at the second option; moving to 20nm Bulk CMOS, this is not an attractive option for many reasons. First, 20nm Bulk CMOS is not a broadly open foundry offering for all customers to access. Even if it was broadly available, the adoption would be limited based on the lack of scaling in power and performance. Additionally, there is no cost per gate advantage in moving from 28nm to 20nm (see figure 1). The lack of scaling for 20nm is driven primarily by the physics of short channel effects and associated negative implications on the transistor characteristics, i.e. leakage and parametric yield, all of which translate into increasing costs. The net result is 20nm Bulk CMOS is not technology that will deliver value to the mobile consumer market.

The third option; jumping from 28nm Bulk CMOS all the way to FinFETS at 16/14nm is fraught with many unknowns and risks in terms of cost and execution. First on the execution risk considerations, this will be the first node the foundry ecosystem is attempting to implement FinFETS, all of which are on aggressive schedules. As we have seen from the experience by Intel in being the first to bring FinFETS to market, they are challenging to execute on schedule in SOC designs. As the industry has observed Intel’s challenges, it is clear to expect the foundry ecosystem to at least be as challenged as Intel has been in bringing FinFETS into production. Once the foundries are able to manufacture FinFETS, there remains the challenge on the design side of moving from a planar transistor structure to a vertical FinFET, which will require full silicon validated design tools and IP. Going from 28nm Bulk CMOS to 16/14nm FinFETS propose the fabless or design houses take two significant jumps simultaneously, skipping a node (20nm) and moving to a completely new transistor structure (planar to vertical). Taking on more than one major technology jump in a given step has always resulted in delays and unanticipated challenges – all of which resulted in missed market windows. Even if the design houses or fabless were confident in their ability to execute both of these challenges simultaneously and factor in the schedule risk on the foundry execution, there remains the challenge on cost. Until FinFETS achieve manufacturing maturity and start reducing process steps and reduce inspection costs, FinFETS will be an expensive technology not suitable for mainstream consumer applications. FinFETS will drive a higher cost per gate (see figure 1) which will translate directly into costs increases at the die level.

Figure 1: Cost Per Million Gates ($) by Technology Note

![Figure 1: Cost Per Million Gates ($) by Technology Note](https://via.placeholder.com/150)

Courtesy of IBS Inc.
The net result is beyond 28nm Bulk CMOS the conventional technology roadmaps are not able to deliver on the technology requirements to meet the needs of the end consumers. 20nm Bulk is not an attractive technology based the lack of scaling and market availability. The move to FinFETs shows long term promise in terms of delivering the power and performance advantages of a fully depleted structure, but there are still a lot of unknowns in terms of cost and time to market risks. The current cost analysis by IBS for both 20nm Bulk and 16/14nm FinFETs represent a challenging set of economics for consumer mobile platforms. Therefore the industry must have a better solution.

As always when the semiconductor industry is faced with a set of challenges or an equation that does not close, the engineering community finds a solution. This is exactly what was done in this case. STMicroelectonics, IBM, Soitec and CEA-Leti partnered to determine a better path for technology beyond 28nm Bulk CMOS. The objectives were clear; develop a technology that continues to deliver scaling in term of power, performance and cost beyond 28nm Bulk CMOS. As is often the case, the simpler the question the more challenging the solution. A solution was found and the results exceeded the objectives. Through cross company collaboration Fully Depleted Planar (FD SOI) technology starting at 28nm was defined, developed and brought to manufacturing. There has been a lot written and demonstrated about FD SOI and its technical advantages of power and performance versus the previously described alternatives, which is broadly accepted in the industry. STMicroelectonics has successfully demonstrated that 28nm FD SOI can deliver performance up to 3Ghz and significantly reduce power without sacrificing performance (see Figure 2). This leaves the remaining question on cost – how does FD SOI compare in terms of wafer and SOC cost versus the alternatives.

IBS has recently published an updated paper that assessed the costs of FD SOI versus the alternative technologies. The analysis by IBS compared the wafer and die cost of 28nm Bulk vs 28nm FD, 20nm Bulk vs 20nm FD and 16nm FinFET vs 14nm FD (see Note 1 about node terminology). Through a deep analysis that was triangulated with multiple sources and factored in all aspects of costs determined the following (Figure 3):

- 28nm FD SOI is cost competitive with 28nm Bulk and delivers a 30% power / performance boost
- 20nm FD SOI delivers a 14% cost reduction vs 20nm Bulk providing an attractive solution for high volume mobile platforms
- 14nm FD delivers a 28% cost reduction vs 16nm FinFET with competitive power / performance.

Based on these findings it is clear that FD SOI provides a significant cost advantage versus all the alternative technologies. When you couple the cost advantage with the ability to have a straightforward design execution roadmap – staying on a planar transistor through 14nm on a superior power and performance technology to Bulk CMOS, FD SOI is a technology that will continue to gain momentum in being adopted. By leveraging FD SOI both the risk of design and the risk of cost are removed from the equation and the path for continued scaling power/ performance/cost is maintained for silicon solutions to continue to meet the accelerating demands of the mobile consumer market.

FD SOI provides fabless and design houses the capability for a straightforward scalable roadmap solution starting at 28nm with FD SOI delivering a significant cost advantage as shown in Figure 4.

**Conclusion**

The technology alternative that represents the best power, performance and cost roadmap forward beyond 28nm Bulk CMOS for high volume consumer application is FD SOI starting at 28nm, delivering scalability for three technology generations. FD SOI provides a path forward with continued scalability, significant cost advantage and execution risk reduction vs all other options. It has been proven by STMicroelectronics public data that FD SOI provides a path for continued technology leadership, with the industries technical community in full agreement of the benefits of moving to a fully depleted structure as early as possible. In addition, the data in terms of cost comparison is compelling at a wafer and even more tilted toward FD when one looks into the true cost to the fabless and
According to Cisco’s Visual Networking Index (VNI), annual global IP traffic is projected to reach nearly a zettabyte (966 Exabyte) by 2015. Video, Mobility and Cloud will dominate the growth. The projected increase of Internet traffic from 2014 to 2015 alone is 200 exabytes, which is greater than the total amount of IP traffic generated globally in 2010. By 2015, one million video minutes which is the equivalent of 674 days, will traverse the internet every second. By 2015, there will be about 15 billion network connections through devices, including machine-to-machine, and more than two connections for each person on earth.

Figure 1: Annual Global IP Traffic Forecast – Cisco VNI

To support this exponential growth, the next generations of networking equipment will offer new levels of packet forwarding rates and bandwidth density. This will spur the need for new generations of packet processors and memory subsystems architecture to support the speed and throughput requirements for packet processing at 400G and beyond. In today’s high-speed networking equipments, like the high-end datacenter switches, edge routers and the core routers, require high-bandwidth, high-access rate, low latency and high density memory solutions. The suitable memory selection for various networking functions is the key to ensure that memory does not become the biggest culprit of data path slowdown. These high packet rates will push against the limits of memory devices – be it DRAM, SRAM, or embedded memory - with packet arrival times becoming much shorter than external memory latency and access times.

An edge or core router used by the service providers in their network consists of Route Processor (RP) Cards and Line Cards (LC). Figure 2 shows a typical edge router with RPs and LCs. There are 2 RP cards in this edge router - one RP is active at all times and the other RP operates in standby mode (used for redundancy). The RP card performs route processing and controls fans, alarms and power supplies in the system. There are multiple Line Cards in a router. Each LC has many ports carrying traffic at specified line rate (e.g. 10G, 40G, 100G, 400G and so on). Each LC implements Layer 1 through Layer 3 functionality consisting of physical layer framers and optics, MAC framing and access control, and packet lookup and forwarding capability. The LC delivers line-rate performance at that specific line rate. While a RP performs routing protocol functions and routing table distribution, the LC forwards the data packets.

Figure 2: Major System Components and Interconnections in the Cisco ASR 9000 Series Routers

ISSI Memory Solution – for 400G+ line rate

Anandarup Bagchi, Director, Strategic Marketing, ISSI
A typical Line Card system utilizes two different types of memory subsystems: a) the packet buffer in the Data Plane where packets are temporarily stored while they await forwarding and b) the table stored in the Control Plane which makes forwarding decisions. Control memories perform functions like packet buffering, table lookup, queue management, and counters/statistics while data memories perform packet buffering. Designers have a variety of high-speed memory to choose from, namely QUAD SRAM, RLDRAM® 2/3, and high-speed DDR3 DRAM. Depending on the particular application and memory access pattern, while taking into consideration memory bandwidth, latency, density, number of devices needed to support the line rate, cost, board area and power, one has to choose the right memory for a given function in a LC.

The packet buffer memory is used to buffer the packets as they are being processed. The packet buffer memory needs to be deep (high density) with high access rate. Packet buffering style accesses are balanced READ and WRITE operations. Packet buffering style accesses are needed for queue buffer function. If cost is the key driver, then DDR3 is preferred over RLDRAM®.

Current memory solutions of high-speed memories, namely QDRII/II+ and DDR4/II+ QUAD SRAM, RLDRAM® 2/3, high-speed DDR3 DRAM provide designers enough alternatives and flexibility to support 10G/40G and in some cases, up to 100G line rate, although 400G and beyond designs are not feasible with current mainstream memory. Figure 5 summarizes the features and functions of memory technologies that are in production today. As line rates increase, the packet rates will increase, and the time between packet arrivals shortens. At 100G, packet arrival rate is 150 million packet per second i.e. packet arrives every 6.4ns resulting in 1.5-2 billion memory access per second (assuming 10-15 memory access per packet). At 400G line rate, the packet arrival rate is 600 million packet per second i.e. packet will arrive every 1.6ns, thus resulting in 6-9 billion memory access per second. As the line rate goes up from 100G to 400G to higher, the number of random memory accesses will go up proportionately. Figure 6 summarizes the random transaction rate or number of memory accesses per second for various control path functions at a target line rate of 400Gbs. At this line rate, a very high random transaction rate memory capable of doing ~10 billion access per second is required to sustain control path functions like statistics and counters, traffic management, Netflow, L2 lookup, wire speed learning, L2/L3 centralized lookup. So, for control path, memory with low latency, i.e. tRC in the range of 10-12ns, density in

<table>
<thead>
<tr>
<th>Parameter</th>
<th>HP-DRAM</th>
<th>RDRAM3</th>
<th>RDRAM2</th>
<th>DDR3</th>
<th>QUAD/DDR SRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage</td>
<td>1.5V-1.8V</td>
<td>1.5V</td>
<td>1.5V</td>
<td>1.5V</td>
<td>1.5V-1.8V</td>
</tr>
<tr>
<td>Speed (MHz)</td>
<td>800</td>
<td>800</td>
<td>1200</td>
<td>1067</td>
<td>550</td>
</tr>
<tr>
<td>Data Inversion</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>Automatic Refresh</td>
<td>y</td>
<td>Y</td>
<td>Y</td>
<td>y</td>
<td>Y</td>
</tr>
<tr>
<td>Overlapped Refresh</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Bus Turnaround Delay</td>
<td>1-2 clock cycles</td>
<td>1-2 clock cycles</td>
<td>1-2 clock cycles</td>
<td>25 clock cycles</td>
<td>N/A</td>
</tr>
<tr>
<td>SRAM Bank</td>
<td>8</td>
<td>16</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Burst Length</td>
<td>2,4</td>
<td>2,4</td>
<td>2,4</td>
<td>2,4</td>
<td>2,4</td>
</tr>
<tr>
<td>Total Memory BW (Gbps)</td>
<td>51.2</td>
<td>51.2</td>
<td>-7</td>
<td>6.85</td>
<td>1.2</td>
</tr>
<tr>
<td>Target Market</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commodity DRAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialty DRAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Performance DRAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Memory subsystem in a typical Line Card

For table lookup or forwarding table, the memory accesses are random. At very high line rate this translates to hundreds of millions of random memory access per second. Low latency is very critical for this type of operation. Lookup style accesses are mostly READs and occasional WRITEs. The QUAD/DDR SRAMs are a good fit for lookup tables which requires very fast random accesses and short read latency. Modern day routing tables have a large number of entries. RLDRAM® with its low-latency, high-density and low $/Gbit cost (as compared to QUAD/DDR SRAMs) is a good choice. High speed DDR3 is an option if cost is sensitive and high density is needed, although the tRC is pretty high (46-52ns) compared to RLDRAM® and QUAD/DDR SRAM. Queue management functions require high bandwidth, low latency, and short bursts. The memory style accesses are random READ and WRITE operations. QUAD/DDR SRAM is a good fit for random READ and WRITE access style. When density is the higher determining factor instead of latency, then RLDRAM® is the preferred choice for queue managers. Counters/Statistics memory is used for billing and other variety of information. These memories perform read-modify-write operation – accesses need to be fast, so low latency is critical. Because of the increasing need to gather more statistics and information, the statistics memory size needed is growing. QUAD/DDR SRAMs are the most popular choice for counters/statistics. Nowadays, RLDRAM® is also used for statistics because of its low latency and high density. If DDR3 is used for counters, then it requires plenty of replication resulting in many DDR3 devices consuming more I/O, more power and more board space.
Take your game changer to the next level.

You didn’t get to this point following someone else. But how do you ensure your breakthrough innovation actually breaks through to the next level? The right guidance can help. Especially if it comes from people who share your passion for changing the game.

You’ll find those kinds of people with the KPMG Technology Innovation Center. Our global network connects you to leading thinkers who can help you identify, evaluate, and derive more value from disruptive technologies. Without disrupting your game.

Join us today.
Visit kpmg.com/techinnovation
In the 1960s, James Early of Bell Labs proposed three-dimensional (3D) structures as a natural evolutionary step for integrated circuits (ICs). Since then many attempts have been made to develop such a technology. Some have even declared the coming decade is the “3D Era” for both transistor shape and layer stacking. However, through silicon via (TSV) connected layer stacking is proving to be expensive, mostly due to the large TSV size and keep-out-zone driven pitch. It is also relatively limited in vertical connection density (about $10^5$/mm$^2$).

The “Holy Grail” of 3D-IC has been monolithic 3D layering. This technology allows a second transistor layer to be constructed directly over the base wafer using ultra-thin (<100nm) silicon, enabling a very rich vertical connectivity. Although it has been difficult to overcome process temperature limitations imposed by the underlying interconnect materials, it now seems that monolithic 3D solutions, in both memory and logic, are becoming practical.

This technology has the potential of becoming the technology that keeps us on track with Moore’s Law, the doubling of the number of on chip transistors that can effectively be produced in volume. With two-dimensional scaling getting harder and costlier and with next generation lithography - Extreme UltraViolet (EUV) - tools continually delayed, monolithic 3D-IC seems to be a very compelling option.

First Adopters

NAND memories drive high volume semiconductor manufacturing today. Samsung announced in August 2013 Mass Production of the industry’s First 3D Vertical NAND Flash. As other memory vendors adopt this path, the International Technology Roadmap for Semiconductors (ITRS) for NAND Flash technology is suggesting that increasing the number of 3D layers per node is the path to future scaling.

A very compelling advantage of scaling in 3D for memory products, as presented in an IEEE 3D-IC conference tutorial, is the ability to process multiple layers simultaneously, supporting higher memory capacity at about a constant cost. This is clearly represented in the acronym BiCS used by monolithic 3D memory innovator Toshiba, standing for Bit Cost Scalable memory.

Further evidence of the viability comes from equipment vendors such as Applied Materials and Lam Research gearing up to support this emerging market. Tools under development, such as Etch and Deposition equipment, allow an increasing number of layers to be processed simultaneously with each new technology node.

Next Adopters

It appears that DRAM vendors are looking for a monolithic 3D path as well. For example, SK Hynix recently licensed BeSang’s monolithic 3D IC technology. Most vendors utilize polysilicon transistors for their 3D NAND transistor stacks, but the leakage of such transistors would be too high for a DRAM implementation. BeSang developed a monolithic 3D technology that utilizes mono-crystalline thin layers, constructing vertical transistors as a second layer on top of a prefabricated device wafer that includes copper interconnect.

The challenge for monolithic 3D technology is the 400°C process temperature limitation imposed by the use of copper (or aluminum) for interconnection. This is the main reason TSV-based 3D IC technology, which allows each wafer to be independently processed, emerged. With this approach, a wafer is thinned, then placed in a 3D configuration, and connected to the substrate with TSV using a low temperature (<400°C) process.

This independent (parallel) processing has its own advantages, however the use of thick layers (>50 µm) greatly limits vertical connectivity as it requires development of new processing flows, and is still too expensive for broad market adoption.

On the other hand, monolithic 3D IC provides 10,000x the vertical connectivity and would bring many additional benefits as recently presented in the IEEE 3D IC conference.

BeSang’s innovative solution is to perform the high temperature process on a donor wafer constructing N-P-N layers. Those layers are transferred on top of the pre-processed base wafer, detaching the donor wafer leaving generic N-P-N layers on top, and completing the processing of vertical transistors.

An alternate innovative flow for monolithic 3DIC was presented at the IEEE 3DIC Conferences of 2012 and 2013. These flows utilize the industry standard Ion-Cut layer transfer technology also known as Smart-Cut®. Ion-Cut has been demonstrated below 400°C and is a volume production qualified process due to its two decades of use for SOI wafer manufacturing. It is estimated that with the reuse of substrates, ion-cut would cost less than $60 per layer.
This alternative flow allows construction of conventional and well understood high performance horizontal transistors. The proven transistor structure, which has been used for many years in DRAM devices, the Recessed Channel Array Transistor (RCAT) is a good example. Figure 1 describes the RCAT process, which constructs the RCAT transistors commonly used in DRAM manufacturing since the 90nm node. The RCAT is quite competitive with standard planar transistors and looks like the inverse of a FinFET.

As can be seen in Figure 2, high temperature dopant activation steps are performed before transferring a bilayer n+/p silicon layer atop Cu/low k using ion-cut. The transferred layers are un-patterned, so that no misalignment issues occur during wafer bonding. After bonding, sub-400°C etch and deposition steps are used to define the transistor. This is enabled by the unique structure of the device. These transistor definition steps directly utilize the alignment marks of the bottom Cu / low k base substrate with conventional nanometer precision, as transferred silicon films are thin (usually sub-100nm) and transparent. Sub-50nm diameter through-layer connections with no dielectric liner needed can be produced due to the thin transistor layer, conventional STI (Shallow Trench Isolation), and excellent alignment.

Recently the industry moved to Hi-K metal gates (HKMG), and later fully adopted the gate last (gate replacement) approach to avoid exposing the Hi-K material, such as hafnium oxide, to high temperature. This could be used for monolithic 3D as illustrated in Figure 2. The first step is to process the dummy gate stack transistors at the conventional high temperature protocol on a donor wafer. The next step is to use ion-cut and a carrier wafer to cause a face-up transfer on top of a base wafer. Finally, perform the gate replacement by removing the H+ damaged oxide and replacing with a HKMG stack using low temperature deposition and etch processes.

It is interesting to point out that just as scaling challenges are mounting, some of the trends associated with dimensional scaling make monolithic 3D easier. For example, the amount of silicon associated with a transistor structure was measured in cubic microns in the early days of the IC industry and has now scaled down to tens of cubic nanometers. The new generation of advanced transistors have thicknesses in nanometers, as seen in the 7nm thin body of Fully Depleted Silicon On Insulator (FD-SOI) transistor recently released to production by ST Micro.

Dimensional scaling has also brought down the amount of time required for transistor activation/annealing, allowing sharper transistor junction definition. In the early days of the industry, annealing and activation took tens of minutes of high temperature furnace based processes. These were replaced later by Rapid Thermal Processing (RTP), which is now being replaced with laser annealing and activation processes. The amount of heat associated with transistor formation has reduced dramatically with scaling, as less volume of silicon gets heated for far less time. Unlike furnace heating or RTP annealing, laser annealing allows the heat to come from the top and directs to only on a small part of the wafer at any given time.

These trends help make it practical to protect the first strata interconnect from the high temperature processing required for the second strata transistor formation. As the high temperature is projected on a small amount of silicon for a very short time and for a small part of the wafer, the total amount of thermal energy required for activation/annealing is now very small.

One of the three most newsworthy topics and papers included in the 2013 IEDM Tip Sheet for the “Advances in CMOS Technology & Future Scaling Possibilities” track was a monolithic 3D chip fabricated using a laser (reported by Solid State magazine “Monolithic 3D chip fabricated without TSVs”). “To build the device layers, the researchers deposited amorphous silicon and crystallized it with laser pulses. The researchers then used a novel low-temperature chemical mechanical planarization (CMP) technique to thin and planarize the silicon, enabling the fabrication of ultrathin, ultraflat devices. The monolithic 3D architecture demonstrated high performance – 3ps compact, energy-efficient mobile products.”
True Circuits introduces a radically new, high performance DDR 4/3 PHY that will change the way you think about DDR.

The True Circuits DDR 4/3 PHY is a high performance, scalable system using a radically new architecture that actively corrects skew among signals and solves most of the problems that plague parallel interfaces. The PHY’s state-of-the-art continuous tuning and automatic training, are the keys to realizing a high performance, low risk DDR system.

Remarkable physical flexibility allows the PHY to adapt to each customer’s die floorplan and package constraints, and is delivered and verified as a single hard macro for easy timing closure with no assembly required.

The PHY is DFI 3.1 compliant, and when combined with the Northwest Logic DDR 4/3 memory controller, a complete and fully-automatic DDR 4/3 system is realized.

Expect more from your DDR interface. Put your trust in our 15 years of timing IP leadership and our deep knowledge of the DDR problem. Celebrate the rebirth of DDR with True Circuits, the timing experts!

WWW.TRUECIRCUITS.COM/DDR_PHY.HTML
Conflict Minerals

Profits from conflict minerals found in the Democratic Republic of the Congo (DRC) have supported conflict, human rights violations and labor and environmental abuses in the region for years. Though the recent surrender of the Congolese Revolutionary Army provides some hope for the future, the group was only recently formed and it is widely recognized that it did not create the conflict minerals problem. 50 armed groups remain, ranging in size from thousands to dozens that are involved in one way or another with conflict minerals. Companies that use these minerals in the design and manufacture of their products and components, called downstream companies, are concerned about these abuses, and are taking action to avoid contributing to conflict in any way. The U.S. Government is helping through legislation and regulations.

Conflict minerals include cassiterite, columbite, tantalite, wolframite and their derivative elements tin (Sn), tantalum (Ta) and tungsten (W) as well as gold (Au), regardless of their source in the world. Collectively these minerals and elements are referred to by the acronym TTTG or 3TG.

The DRC is the size of Texas and Alaska combined with a population of 75 million who are among the poorest people in the world. The DRC has mineral reserves of $24 trillion, the richest of any country in the world. The country has been in a civil war for 15 years and millions of people have been killed during that time. The death rate remains very high due to violence and the sickness and starvation exacerbated by it.

Legislation

A rider (Title XV, Section 1502) was attached to the U.S. Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010. The bill required that U.S. public companies provide a report if any of their products contain conflict minerals that were sourced in the DRC or adjoining countries. It also stated that the company had responsibilities to determine the source and chain of custody of those minerals and that there must be an independent private sector audit of the report.

It was determined that this requirement would be administered by the U.S. Securities and Exchange Commission (SEC). The SEC developed policies and procedures to administer Section 1502. There were hearings and modifications and the Final Ruling was passed by the SEC on August 22, 2012 and issued as a 356 page document. Shortly afterwards, the U.S. Chamber of Commerce and the National Association of Manufacturers filed suit to “modify or set aside in whole or in part” the SEC rules governing so-called conflict minerals. In a July 23, 2013 opinion, the U.S. District Court for the District of Columbia upheld the SEC’s rule requiring disclosure of companies’ use of conflict minerals originating in and around the DRC.

Filing Requirements

The first filing year is 2013, beginning on January 1, 2013 and ending on December 31, 2013. The first filing is due on or before May 31, 2014. The requirement is ongoing with reports due on May 31 after the end of each compliance year. There is a reasonable expectation that additional materials from the same or other regions may be added to the requirements. The conflict minerals and covered countries in the rule align with those identified by the U.S. State Department (available on the State Department website). If the State Department modifies its list of conflict minerals or covered countries, the SEC rule automatically follows suit.

The SEC Final Rule applies to all SEC “issuers”, including foreign issuers. If a company is not public, does not manufacture products or manufactures products that do not include conflict minerals, they need not file with the SEC. If a public company produces products that include conflict minerals, a determination must be made as to whether any of these minerals were sourced from the Conflict Countries (the DRC plus the 9 adjoining states). This process is referred to as a Reasonable Country of Origin Inquiry (RCOI) and must be performed in accordance with an internationally recognized framework, such as the OECD Guidelines for Multinational Enterprises. If there is reasonable doubt as to the source of the conflict minerals, a more detailed due diligence effort must be undertaken to determine the source with as much specificity as possible.

Public companies must “file” (not “furnish”) with the SEC, which falls under Article 18 of the SEC code. Filing has much broader liabilities for not filing or for providing incorrect information including the potential for fines and incarceration. A public company does NOT have to become conflict free, but they must follow the procedures outlined in the SEC Final Rule. A private company, though they need not file with the SEC, can be held liable if they provide false information. The company must provide both the procedure and the results of the RCOI and due diligence work.

Unless the determination of the process is that the products are conflict undeterminable, an independent third party audit must be carried out and the report filed with the other SEC filings.

Though the RCOI and due diligence work is very much supply chain oriented, the actual SEC filing is by product or product group. The statute says that all products are to be reported, but not all
companies in the supply chain must be reported. This gives some flexibility in how many players are reported.

On the surface, it would appear that private companies in the semiconductor supply chain do not get involved, but this is far from the truth. As public companies, semiconductor or otherwise, must trace their supply chain down to the Smelter or Refiner (SoR) level, they must “pass through” private companies in the chain. While there is no legal requirement to participate, there is pressure from upstream customers to support the conflict minerals compliance effort. They need to do the RCOI and due diligence work, they differenc is that they do not need to file with the SEC.

**Conflict Minerals in the Semiconductor Industry**

For an integrated circuit (IC) or a discrete semiconductor device to be conflict free, both the chip and the package must be conflict free. If the die is conflict free and the package is not, the IC is not conflict free. There is a reasonable probability that a simple IC chip with but a few levels may not contain any conflict minerals. As the chip gets more complex with more levels, the probability goes up that tungsten, gold or tantalum may become part of the chip. Package assembly has a generally higher probability of containing conflict minerals due to materials such as gold bond wire, tin content in lead finish or solder balls for package to board connections. It is highly unlikely that an IC of any significant complexity will be free of all conflict minerals.

**Semiconductor Supply Chain Structure**

First, we have a semi-fixed set of resources that produce die and a semi-fixed set of resources that do packaging. A few fabs (captive or foundry) are built and shut down each year, but there is stability and their identity is known on a worldwide basis. The same is true for assembly operations (captive or OSAT). This means that virtually all IC’s are manufactured by a known set of factories. This is far different than products, like computers or airplanes, that have many levels in their bill of material and whose supply chain includes myriad manufacturers or fabricators around the world. Fabless companies have about 75 foundry and 75 OSAT companies from which to choose. This means that the potential exists for information collected about Foundry X or OSAT Y to be shared by many fabless companies.

Second, the direct materials used in semiconductor manufacturing are typically very high purity. In fab, WF6 process gas or TiW sputtering targets are used and are highly controlled for content and purity. In assembly, gold bond wires have tightly controlled composition. Even Tin, used in lead finish or package connection, is controlled for contamination and consistency. The use of high purity direct materials puts the industry in close proximity to the smelters and refiners that must be tracked.

**What needs to be done?**

Many companies believe that conflict minerals compliance consists of nothing more than gathering EICC-GeSI forms from first tier suppliers. This is merely the tip of the iceberg. The RCOI in the SEC Rule requires that companies track the conflict minerals to the smelter or refiner. Depending on the RCOI outcome, a much more in-depth due diligence analysis may be required. The outcome of these processes will determine whether a third party audit is required. Public companies must file these documents with the SEC. Private or public companies need to provide this information to their customers who must include it in their filing. The SEC stipulates that an internationally accepted standard be used for RCOI and due diligence processes.

**Typical Company Compliance Flow**

- Form a compliance team and draft a Conflict Minerals Corporate Policy statement
- Put a conflict minerals management system in place. This consists of policies and procedures for conflict minerals management. It also includes development of a system (typically using Excel or Access) to provide a framework to track info from direct suppliers as well as tier 2, tier 3, etc. to the SoR level.
- Get up to speed on all conflict minerals compliance requirements.
- Gather information from tier 1 suppliers, then parse and store information, compliance documents, expiration dates, contact information, tier 2 suppliers used, etc.
- Gather information from tier 2 suppliers, then parse and store information, compliance documents, expiration dates, contact information, tier 3 suppliers used, etc.
- Gather information from tier 3 . . . and beyond to reach the SoR level.
- As information is received, begin the RCOI analysis and due diligence process as required, storing not only the results of what is found, but also information as to the RCOI and due diligence process utilized, as this must also be reported to the SEC by public companies.
- As determination is made of probable compliance and non-compliance, begin development of a product oriented strategy on how to report to maximize the percentage of products that are compliant.
- Have audit performed, if required.
- If public, generate forms SD and Conflict Minerals Report (CMR) and submit to SEC with audit report.
- If private or public, make information available to customers that require it.

**Alternate approach**

Conflict minerals reporting is an ongoing requirement that will likely grow in scope over time. It is not a core competency and does not provide a value add to semiconductor companies at any level. The work may be performed by each individual company with great duplication of effort, or performed by specialists to reduce redundancy.

The unique nature of the IC supply chain and bill-of-materials structure enables the semiconductor industry to be leaders in the efficiency of conflict minerals compliance. This is good not only for our industry, but for the greater electronics supply chain.

**About the Author**

Ron Jones is CEO of N-Able Group International; a semiconductor focused consulting and recruiting company. N-Able Group utilizes deep semi supply chain knowledge and a powerful cloud based software application to provide Conflict Mineral Compliance support services to companies throughout the semiconductor supply chain including fabless, foundry, OSAT and materials suppliers. ron. jonei@n-ablegroup.com.
Manufacturing for a Smart World

TowerJazz manufactures integrated circuits with geometries ranging from 1.0 to 0.13-micron for more than 200 customers worldwide. The company specializes in customized analog solutions for differentiated products in various markets including radio frequency, high-performance analog, power, imaging, consumer, automotive, medical, industrial and aerospace & defense.

- High Performance Analog (HPA) Radio Frequency (RF) CMOS SOI
- Integrated Power Management (PM) Platforms
- CMOS Image Sensor (CIS) Technology
- Mixed-Signal CMOS Platform
- Micro-Electro-Mechanical Systems (MEMS)
- Worldwide Design Enablement Tools and Models
- Transfer Optimization and development Process Services (TOPS™)
- Aerospace and Defense (A&D)
The data center houses critical network systems are vital to the daily operations of modern organizations. However, with rapidly rising maintenance costs for the most up-to-date technologies to operate and regulate the data center, many companies continue using outdated methods. Refusal, or perhaps inability, to embrace innovation in data center solutions is shortsighted and inevitably leads to costly problems in terms of money and time. The most pervasive and arguably most threatening issue is the overall lack of efficiency since it correlates with many other issues. Implementing a solution for monitoring and managing the data center, such as Data Center Infrastructure Management (DCIM) tools, reduces power consumption by 15-25 percent1. Affirmative action’s increasing energy efficiency shed a positive light on data centers, which in this time of green energy enthusiasm, can easily be seen as a necessary evil. Exploration for innovative data center solutions must continue. These solutions not only treat issues after they occur but also solutions that are preventative and proactive.

**Data Center Management and Efficiency**

Efficiency is the ideal organizations strive for in all aspects of their operation. Inefficiency, which is a high quality final result notwithstanding, wastes valuable resources (e.g. time, money). These losses are not trivial and are ignored at the detriment of the final product. It comes as no surprise; increasing efficiency is the battle cry for data center operators.

Similarly, information within a data center must travel through the network and reach the intended destination with low latency and high accuracy. Lowering the power consumption and improving the throughput within the data center grows increasingly more difficult as data centers continue growing tremendously.

**Software-Defined Networking**

Software-Defined Networking (SDN) has become a solution that many operators are turning to in hopes of building a more organized and intelligent data center. Google made waves just last year when they announced their use of a SDN-based system called OpenFlow, which not only improved their efficiency, but also provided them with huge savings2. OpenFlow separates the control panel and data plane in a network device giving the operator increased control. For example, think about data as taxi cabs. Using the old system, as the cab travels to its destination, it may run into traffic, bad weather, or an accident. At the time it runs into each situation, it changes routes, which could take much longer than originally planned to reach its end point.

[With the implementation of OpenFlow], all cars are autonomous and can report their whereabouts and plans to some central repository that also knows of weather conditions and aggregate traffic information. Such a system doesn’t need independent taxi drivers, because the system knows where the quickest routes are and what streets are blocked, and can set an ideal route from the outset.

Although more manual involvement is necessary with OpenFlow as opposed to traditional data center structures, the ability to optimize the system through path determination is a strong incentive.

**Integrated Systems**

Another solution aims to re-imagine the most basic building blocks of data centers: hardware. Optimizing the hardware alongside software realizes tangible benefits where previously, only potential existed. In a report generated by Gartner, hardware integrated systems are defined as “a class of data center systems that deliver a combination of server, shared-storage and network devices in a preintegrated stack”1. Exactly as the name suggests, these systems are preintegrated to improve flexibility, time to solution, and reliability, with an added bonus of cost per user savings. However, even with these significant claims, the challenge remains of convincing organizations to shoulder the upfront investment to transition their systems without a quantifiable value proposition. Although integrated systems revenue only accounted for 3.5 percent of the data center hardware total in 2011, major players continue to perfect integrated systems (e.g. Cisco, Dell, IBM, Oracle, HP)3.

**Data Center Infrastructure Management (DCIM)**

While the tools mentioned above seem to be solid solutions to combat inefficiency, they would be pointless without high-level tools monitoring and managing all the systems and applications. For this reason, more and more organizations are turning to DCIM tools, as they allow for centralized management of a data center. However, as simple as that definition sounds, there are many facets to DCIM, which cover both the physical and virtual infrastructures of data centers.
DCIM offers real-time reporting as to what is happening in the data center. It is able to give a broad, big-picture view of the operation and, with just a few simple clicks, a more specific view of events that might need attention. Additionally, DCIM sends alarms/alerts when an error has occurred, thereby improving the response in terms of time needed to locate the problem area as well as knowing right when a threat occurs. Real-time reporting provides greater visibility into the system, which in turn produces many other benefits for data center management.

**Energy Consumption**

Another benefit of implementing DCIM tools is the ability to track energy usage of a data center. Since data centers are built to handle potential peak loads, there are substantial inefficiencies involved during non-peak times. This strategy is known as the capacity safety gap or over-provisioning. DCIM tools map a trend of energy consumption over a period of time in order to make adjustments based on those trends. By implementing DCIM, data centers reduce power consumption by at least 15-25 percent.

**Liquid Cooling**

Cooling is another issue directly related to increased energy usage in data centers. According to research, data centers are usually overcooled, wasting as much as 45 percent of energy costs. Not only are costs wasted, but also air cooling, the most commonly used today, is extremely inefficient since it requires a considerable amount of energy on top of the considerable amount of energy already consumed by the compute and storage operations in a data center. One solution scientists are looking into is liquid cooling. Water is a significantly more efficient medium than air for heat removal. During tests conducted over a span of six months at Lawrence Berkeley National Laboratory, liquid cooling provided an estimated 16 to 20 percent energy reduction in a server load of 50 percent. Although liquid cooling has been deployed in supercomputing sites, it is hard to tell when, or if, smaller facilities will adopt liquid cooling as their primary method of heat removal. One thing remains clear, though; better solutions in cooling are required to maintain acceptable energy usage without large amounts of waste.

**Renewable Energy**

From a green perspective, energy waste widens the carbon footprint of the data center. As data centers are critical to most organizations, it is not realistic to expect them to disappear or fundamentally change any time in the near future. As organizations, and even individuals, realize the importance of minimizing their carbon footprint, renewable energy is a subject matter that should be discussed. If a necessary evil of data centers is their avarice for electricity, at least in the foreseeable future, shouldn’t considerable time be spent on renewable energy research? Buying wind power is a solution used by both Google and Microsoft. While wind power is a great source of renewable energy, many factors must be taken into consideration. In Microsoft’s case, the wind farm to supply their wind power is not built yet. Construction starts in 2014 with power generation starting in 2015. Would this wind farm have been built anyway if not for Microsoft? According to Stanford research fellow and expert on data center energy consumption Jon Koomey, “making sure the renewable energy project you are investing in is additional is the way it should be done.” Wind power may be a great alternative to traditional sources of energy for large-scale companies like Microsoft and Google. However, more time and additional research is required before renewable energy, like wind power, becomes a truly viable option for the masses.

**Disaggregation**

Since most companies use data centers, not just Microsoft or Google, increased emphasis needs to be placed on making data center improvement and optimization more affordable. At Facebook’s Open Compute Project summit in January of this year, members pushed for disaggregation as the next step in data center innovation. If implemented successfully, disaggregation separates the core components of system design, allowing for flexibility in upgrading independent parts. The benefits of disaggregation are obvious to most data center operators, who in the past have been reluctant to upgrade their systems because of the difficulty incurred from switching out mutually dependent components. Historically, upgrading necessitated large capital expenditures that smaller organizations could not afford, yet with disaggregation; specific parts can be upgraded independently translating to lower change costs. While the Open Compute Project may become an intuitive and obvious step towards improved hardware migration within data centers, not everyone agrees. In fact, some systems vendors whose business model depends on the existing infrastructure are hesitant to cooperate with the movement toward disaggregation. Currently, those vendors exert control by making their parts highly dependent on each other. This control translates into higher profits for the vendor since multiple parts must be upgraded simultaneously. In order to successfully integrate disaggregation as the norm for data center components, systems vendors must be forced to realize the value in interoperable parts. Alternatively, new system vendors can produce equipment based on open architecture principles.

**Conclusion**

As the number and size of data centers constantly grows, they must be continually studied, re-imagined, and innovated. As something so vital to the operation and maintenance of organizations everywhere, there is no excuse for data centers to continue to be so inefficient and energy draining. As researchers and scientists discover new and exciting ways to solve issues plaguing data centers, it’s no use if the organizations themselves do not realize the need for revitalization. Data center operators must educate others within the organization that changes at the core, the building blocks of the data center, are necessary and worthwhile. Hardware and software infrastructures and management tools all need to work together in order to create the most valuable, efficient data center. Surface-level fixes may seem good enough, but the importance of the data center calls for deeper and perhaps even disruptive solutions. If continuous progress is made on innovative answers to data center issues, organizations can rest easy (maybe just easier) knowing that their future data centers will benefit from solutions making them more intuitive to manage, increasingly energy efficient, and based on open architecture principles.

**About the Authors**

Ariel Ueng is a marketing associate at SKTA Innopartners. She holds a Bachelor of Arts degree in Media Studies from the University of California, Berkeley.

Angel Ornontia is the Business Development Director at SKTA Innopartners. A veteran of the semiconductor and hardware industry in diverse roles including IC design, technical lead for implanted medical device development, business development, corporate law, and IP litigation. Angel is now focused on
Comprehensive Foundry Solutions

UMC’s collaborative foundry approach is based on strengthening our partnerships with customers, providing competitive advanced technology, and committing sufficient capacity to secure mutual long-term growth opportunities. We offer advanced technology down to 28nm, multiple specialized technologies, and a broad IP portfolio including ARM, Faraday, and Synopsys offerings. These foundry solutions establish a comprehensive process platform to streamline our customers’ path to SoC silicon success, for now and the future.
The OpenCL provides many key characteristics that allow efficient FPGA applications to be written in a standard multi core programming language. OpenCL is a language and a set of APIs that allows for the expression of algorithms on heterogeneous platforms. It is based on C99 and contains keyword extensions that allow a user to specify parallelism and memory hierarchy. Parallelism is a key factor for high performance FPGA applications. Typical designs implemented on the FPGA’s programmable fabric operate in the range of 300-500 MHz. While these clock speeds are lower than the multi GHz range that is typical for high-end microprocessors, the programmable fabric of the FPGA compensates with extreme amounts of parallelism as each LE is essentially able to operate independently. Similarly, explicit control of the memory hierarchy is a key factor in the performance of the overall system. It is not enough to create an efficient data-path, but we must ensure that it can be fed with data. For example, high end FPGA devices may have an external memory bandwidth of approximately 80-100 GB/s; however, the internal memory blocks have an aggregate bandwidth approaching 5-6 TB/s. Thus, organizing algorithms such that windows or blocks of data are brought from external to internal memory before computational processing begins is a very effective way of ensuring high performance. OpenCL provides very natural mechanisms for describing algorithms that are highly parallel in nature and require careful management of I/O bandwidth to obtain high performance.

Another key aspect of the OpenCL standard is its heterogeneous nature. Programs are written entirely in software; however, key functions, known as OpenCL kernels, can be offloaded onto different portions of the heterogeneous platform. When these functions are targeted at the programmable fabric, Altera has created an automated compilation flow which converts the software representation of these functions into a fully customized data-path that is dedicated to implementing that particular OpenCL kernel function. There is no longer a need to describe these data-paths using complex and tedious HDL code. The programmer can stay in a software design environment without needing to venture into the world of hardware design. This methodology now allows a programmer to seamlessly manage the partitioning of algorithms on modern heterogeneous FPGAs. For example, there are often portions of an application such as control processing that are better suited for implementation on a hardened processor system rather than programmable logic. Similarly, portions of applications that are doing large amounts of parallel number crunching such as filters or spectral analysis are typically better suited to the programmable fabric. By keeping the entire algorithm in a single software specification, the programmer can now easily explore these different choices simply by targeting their kernels at a different portion of the heterogeneous chip. It is the compilation tool that will automatically take care of the underlying implementation. Using the same software source code, a binary executable for a processor system can be created or a customized implementation of data-path for FPGA fabric can be generated depending on the target. This ability frees the programmer from having to manually create different implementations to evaluate system partitioning decisions.

About the Author
Desh Singhal is a Director of Software Engineering at Altera’s Toronto Technology Center. Desh’s team is responsible for Altera’s high level tool solutions including DSP Builder and Altera’s SDK for OpenCL. His charter is to develop high level design tools which allow designers to create applications for FPGAs with a higher level of productivity than traditionally possible. Previously, his group was responsible for a number of optimization algorithms in Altera’s Quartus II CAD tool. These include Logic Synthesis, Line-level incremental compilation, Physical Synthesis, Metastability Analysis, and IP core optimizations. Desh holds a PhD from the University of Toronto in the area of timing closure techniques for high speed FPGA designs and has authored over 50 patents and publications on FPGA technology.

References

Innovative Solutions continued from page 24

reinvigorating the funding ecosystem for semiconductor and hardware to foster innovation in the Silicon Valley.

5 In 2013, the San Francisco Board of Supervisors passed a measure to restrict the building of new data centers in San Francisco to zero carbon footprint facilities. This measure was passed by a 7-4 vote and was supported by a coalition of environmental groups and San Francisco residents. The measure was supported by a coalition of environmental groups and San Francisco residents. The measure was intended to reduce the carbon footprint of data centers and to encourage the adoption of renewable energy sources.
6 In 2013, the San Francisco Board of Supervisors passed a measure to restrict the building of new data centers in San Francisco to zero carbon footprint facilities. This measure was passed by a 7-4 vote and was supported by a coalition of environmental groups and San Francisco residents. The measure was supported by a coalition of environmental groups and San Francisco residents. The measure was intended to reduce the carbon footprint of data centers and to encourage the adoption of renewable energy sources.

7 In 2013, the San Francisco Board of Supervisors passed a measure to restrict the building of new data centers in San Francisco to zero carbon footprint facilities. This measure was passed by a 7-4 vote and was supported by a coalition of environmental groups and San Francisco residents. The measure was supported by a coalition of environmental groups and San Francisco residents. The measure was intended to reduce the carbon footprint of data centers and to encourage the adoption of renewable energy sources.

8 In 2013, the San Francisco Board of Supervisors passed a measure to restrict the building of new data centers in San Francisco to zero carbon footprint facilities. This measure was passed by a 7-4 vote and was supported by a coalition of environmental groups and San Francisco residents. The measure was supported by a coalition of environmental groups and San Francisco residents. The measure was intended to reduce the carbon footprint of data centers and to encourage the adoption of renewable energy sources.

9 In 2013, the San Francisco Board of Supervisors passed a measure to restrict the building of new data centers in San Francisco to zero carbon footprint facilities. This measure was passed by a 7-4 vote and was supported by a coalition of environmental groups and San Francisco residents. The measure was supported by a coalition of environmental groups and San Francisco residents. The measure was intended to reduce the carbon footprint of data centers and to encourage the adoption of renewable energy sources.
EMERGENCE OF ALTERNATIVE MANUFACTURING MODELS

The manufacturing commitment curve (MCC) shown in Figure 5 examines the continuum of possible manufacturing structures in the context of the amount of invested capital and the corresponding control over capacity. The MCC is key in understanding the ways semiconductor firms will address manufacturing concerns going forward.

At the low end of the MCC is the classic fabless model that requires little invested capital, but also limits the degree of control over manufacturing and guaranteed capacity. The classic IDM model – found on the opposite end of the MCC – does provide a large degree of control over processes and capacity, but that freedom comes at the cost of high invested capital. As semiconductor companies strive to avoid the sub-optimal scenarios that develop as a result of a dominant foundry player squeezing the profit pool while also addressing the rapidly rising costs of new fab development, alternative manufacturing models are likely to emerge in the chasm between the classic fabless and the classic IDM models on the MCC.

Two frameworks are especially likely to become prevalent: Equity for capacity. In this model, a semiconductor firm or OEM makes an equity investment in a foundry or fab operator. This equity investment gives the investor a stake in the overall success of the fab operations (an undivided interest in the entire manufacturing operation), provides insight into the financial state of the fab operator, and ensures guaranteed access to capacity while minimizing the total capital investment. In turn, the fab operator gets an infusion of cash and a long-term partner/customer.

Cooperative (co-op). Under this arrangement, a fab is co-owned and co-operated among multiple semiconductor firms. The ownership may be a portion of capacity or module, with expenses shared between partners. While this model has been tried with varying degrees of success, firms have learned to avoid some of the pitfalls of such structures and are likely to bring about additional co-op deals in the next few years.

MOVING FORWARD

The semiconductor industry is entering a new era. Rising costs, slowing growth, and increased concentration in high-growth markets, combined with an unstable foundry model, are fundamental forces that will drive its near-term evolution and lead to the creation of alternative manufacturing models. These models will attempt to marry the low commitment of a fabless model with the increased control of an IDM framework. Beyond simply keeping an eye on competitors, the semiconductor segment will need to be mindful of the entire ecosystem and the intricate dynamics of the supply chain in order to proactively define its optimal future rather than become an impotent participant in it. Change will be difficult. The firms that best adapt to the pressures of unlocking novel operating models and partnerships have an opportunity to maximize their potential and lead the wave of innovation that will define the 21st century.

About the Author

Stephen Rothrock founded ATREG in 1998 in response to the global semiconductor industry’s need for strategic advisory services related to the monetization and disposition of complex and formerly illiquid assets. After initially focusing on buying and selling front-end and A/T wafer fabs, the company quickly expanded its offering to include the transfer of business units, intellectual property, long-term supply agreements, and complete tool lines. Over the past three years, ATREG has added new consulting services to its core business of fab sale transactions in the realms of manufacturing strategy, intellectual property licensing, bankruptcy liquidations, strategic capital raising, and the establishment of joint-venture manufacturing partnerships.

In 2012-2013, Stephen spent 18 months in Japan to set up ATREG’s Tokyo office and oversee the firm’s Asian operations. Prior to ATREG, he established Colliers International’s Global Corporate Services initiative and headed the company’s U.S. division based in Seattle, WA. Before that, he worked as Director for Savills International in London, UK, also serving on the UK-listed property company’s board. Stephen spent four years in Paris, France working for an international NGO. He holds an MA degree in Political Theology from the University of Hull, UK and a BA degree in Business Commerce from the University of Washington in Seattle.

Monolithic 3D IC continued from page 18

thermal anneal to boost performance of 3D memory devices.

In conclusion, while dimensional scaling is becoming more difficult, it allows Monolithic 3D to become easier. The industry should be able to keep scaling one way or the other, potentially even both, while continuing to enjoy the benefits.

About the Author

Zvi Or-Bach is the founder President and CEO of MonolithIC 3D™ Inc. Or-Bach has a history of innovative development including the breakthrough of monolithic 3D IC and fast-turn ASICs for over 30 years. Prior to MonolithIC 3D, Or-Bach founded eASIC in 1999 and served as the company’s CEO for six years. Earlier, Or-Bach founded Chip Express in 1989 (recently acquired by Gigopix) and served as the company’s President and CEO for almost 10 years, bringing the company to $40M revenue. He holds over 100 issued patents, primarily in the field of 3D integrated circuits and semi-custom chip architectures. He is the Chairman of the Board for Zeno Semiconductors, Bioaxial and VisuMenu.

Note: smart-cut® is a registered Trademark of Soitec.
order of 1Gbit to 4Gbit, and able to support ~10 billion accesses per second would be a perfect candidate to sustain 400G line rate control path functions.

**Figure 6: Memory accesses for various control path functions at 400Gbps**

For packet buffer memory, a raw bandwidth of 3X-4X line rate is needed. To sustain 400G line rate, the packet buffer memory bandwidth should be in the range of 1.2 - 1.6 Tbps. The packet buffer also needs to store 10ms to 150ms of data due to roundtrip latency. At 400G line rate, it translates to 4Gbit for a shallow buffer or up to 64Gbit for deep packet buffer, depending on the application buffering requirement.

Recently, there has been introduction of new memory technologies; the Hybrid Memory Cube (HMC) and High Bandwidth Memory (HBM) are among the notable ones.

HMC was aimed at solving the classic “memory-wall” problem in high-performance computing – providing a high performance memory with lower power per bit in a small footprint. The HMC architecture uses a small, high-speed control logic die that sits beneath vertical stacks of up to eight DRAM die, all connected with through-silicon vias (TSV). The DRAM uses the traditional DRAM core cell architecture, but it is restructured so that it uses memory vaults instead of arrays. In a HMC, there are 16 independent vaults which can be thought of as channels. A high-speed SerDes interface connects the memory cube and the host processor (or other memory cubes) using 16 transmit lanes and 16 receive lanes, each running at 10Gbs. Although it is currently being considered for networking applications, some challenges still remain for it to be a perfect solution. The serial links instead of parallel links definitely provide the advantage of much fewer pins on the host ASIC/processor to connect to the memory cube. However, the serial interface introduces longer round-trip latency. The tRC is traditional DRAM like, but only slightly better statistically (35~40ns) because of the huge parallelism due to multiple vaults per die and multiple die.

HBM DRAM is a JEDEC initiative - wide I/O stacked DRAM with TSV. The HBM DRAM is optimized for high-bandwidth operation of a stack of multiple DRAM devices across a number of independent interfaces (channels). Each DRAM stack will support up to 8 channels and each channel interface maintains a 128b data bus operating at DDR data rates. The 1024b wide interface between the host and the memory provides high-speed, low-power operation. Each die contributes additional capacity, and bandwidth is decoupled from density. JEDEC HBM’s focus in the graphics market left a huge gap in the market for networking applications, where product lifecycle is much longer than Graphics ASSP. The table here compares various memories for networking buffering application. Stacked HBM-like memory, running at 600-800Mhz (i.e. bandwidth of 1.2Tb/s - 1.6Tb/s), and 40ns tRC can be a very good option for buffer memory. In terms of bandwidth per watt, bandwidth per dollar and density per dollar, stacked HBM is a clear winner. For the control memory, RLDDRAM-like memory with HBM interface and a capacity of 1-4Gb would be a very good fit. In other words, a wide I/O high bandwidth memory, with low random-bank cycle (tRC-10ns), multibank write, minimal bus turnaround delay, access granularity of 64b to 256b, no tFAW or tRRD, no tWTR, and unlike DRAM that does not need an ACTIVATE, READ, and PRECHARGE, will be able to meet the memory access rate of ~10 billion access per second for 400G line rate.

**Figure 7: Comparison of Parallel, Serial and Stacked Memory**

The trend in the industry is leaning towards multi-die package solution using 2.5D and 3D TSV integration. The key driving factors are bandwidth, latency, power, form factor and cost. In High End Networking Devices I/O Switching typically consumes 20 – 40 percent of device power. Multi Die Packaging reduces I/O power through reduced path length, i.e. reduced capacitive loading and the ability to switch with reduced drive strength, i.e. lower voltage switching. Multi die options configurations are typically more expensive than single package die options however overall cost of ownership may be lower once board area or product size is considered.

If the HBM with RLDDRAM-like memory die, can be integrated with the ASIC/NPU/FPGA using 2.5D integration on an organic/silicon interposer, there will be a significant improvement in power, board area, and latency. Moreover, if the memory die are stacked (3D integration) to provide a capacity of 32Gb to 64Gb, then one universal memory can be used for both control and data memory requirement for 400G and beyond line rate requirements.

In summary, requiring a high performance (bandwidth), low latency, low power per bit, very high random transaction rate, yet space and cost-efficient memory solutions to support very advanced networking applications of 400G and beyond line rate has posed quite huge challenge to the memory industry. The current available solutions and emerging technologies have their own advantages and disadvantages. From what has been discussed so far, it appears that RLDDRAM-like 3D TSV stacked DRAM die with HBM-like interface, integrated with the host using 2.5D TSV would be an ideal solution for high-end networking line card.

**About the Author**

Anandarup Bagchi [anand_bagchi@issi.com] is Director, Strategic Marketing at ISSI and is responsible for the Communication and Networking vertical market. Prior to ISSI, Anandarup spent many years at Cisco in business and technology leadership roles across multiple organizations. He has also been involved with several technology startup companies through various stages of their evolution. Earlier in his career, he has worked at Sun Microsystems and National Semiconductor. Anandarup holds a BS in Electronics & Electrical Communication Engineering from Indian Institute of Technology (IIT), Kharagpur and a MBA from Carnegie Mellon University.
or .18 micron technology which allows fabs to continue to utilize and depreciate their older lines and receive a better cost structure. Actually, this was one of the early key enablers of MEMS as it was a way for manufacturers to extend their fabs and back-end.

Q: Another common topic at GSA is More than Moore (MtM). How do you extend the same design principles that have driven digital device scaling to analog circuitry and how do we integrate these technologies within a SoC. How do you see this playing out in the MEMS market?

A: Certainly we see integration of various technologies within the same package continuing as referenced above, but certainly not to the same degree as in the digital world; this won’t happen unless one of the major foundries comes up with a new technology platform that provides standardization across the MEMS industry. Still, MEMS is a burgeoning market and there is quite a bit of architectural innovation left, so efficiencies will definitely continue to improve through heterogeneous integration.

Q: At the 2010 MEMS Technology Summit Conference at Stanford University, Muenzel Horst, now President of Bosch Akustica, presented a vision for 7 trillion “sensory swarms” by 2017. Do you agree with this view? What is needed to improve the MEMS ecosystem to support the Trillion Sensor vision?

A: I’m not sure how others are extrapolating, but my opinion is we currently aren’t on track. There’s no doubt that we’ll get there as sensors are becoming ubiquitous, especially with the proliferation of the IoT, but I don’t see it happening by 2017.

Q: What is needed to improve the MEMS ecosystem to support the Trillion Sensor vision?

A: MEMS technology needs to be one or two orders of magnitude more cost effective; the cost structure (and reliability) needs to support that.

The signal processing and algorithms need to improve. In fact, there is currently a lot of innovation being done in this area. Probably most important is low-power connectivity. There has to be a viable mesh network that allows the sensors to talk to one another at the local level (what UCB calls the Swarm and Cisco calls the Fog) and then connect to the cloud.

Q: Closely associated with the Trillion Sensor vision is the Internet of Things (IoT). What are the growth markets associated with IoT and what does IoT mean for InvenSense’s future?

A: If one assumes wearables as a subset of the IoT and within this subset the various categories such as health, fitness, computing, and immersive gaming, it is easy to see a common thread amongst most of these categories which I consider the killer app, motion, which is where we benefit. For example, if one looks at the conceivable number of sensors that can be attached to a person for health and fitness to measure such things as their heart rate, glucose, and movement; or the number of sensors that can be used in wearableoggles and such for a completely immersive and virtual gaming experience. If we look at these applications, we see that the wearable market can easily exceed tens of billions of units. And then if we apply the same math to other devices within the IoT spectrum that require sensing, monitoring, and controlling, the numbers are staggering. Of course for all of this to become a reality, several things need to occur in the ecosystem, but we do see the opportunity.

Q: InvenSense recently joined GSA. What was the driving factor and what has been the greatest benefit to date?

A: GSA gives me access to data and information that provides insight into what is going on in the industry. GSA also provides me tremendous networking opportunities. Sharing awareness of policies and the strategy of dealing with these new policies is a very important aspect to me. It allows members to gain insight into the IoT spectrum that require sensing, monitoring, and controlling, the numbers are staggering. Of course for all of this to become a reality, several things need to occur in the ecosystem, but we do see the opportunity.

Note 1

IBS is using the term “20nm FD SOI” for the generation that follows 28nm FD SOI – as it indeed uses the same back-end of line (BEOL) structure as 20nm bulk CMOS technology – while STMicroelectronics calls it “14FD”, based on the fact that its power and performance merits allow to position it as an alternative not only to 20nm planar bulk growth markets associated with IoT, but also to 16nm/14nm FinFETS – and the latter already use a 20nm-like BEOL, too.

Similarly, IBS’s “14nm FD SOI” denomination would correspond to STMicroelectronics’s “10FD”.

About the Author

Christophe Maleville has been Senior Vice President of Soitec’s Microelectronics BU since 2010. He joined Soitec in 1993 and was a driving force behind the company’s joint research activities with CEA-Leti. For several years, he led new SOI process development, oversaw SOI technology transfer from R&D to production, and managed customer certifications. He also served as Vice President, SOI Products Platform at Soitec, working closely with key customers worldwide. Christophe Maleville has authored or co-authored more than 30 papers and also holds some 30 patents. He has a PhD in microelectronics from Grenoble Institute of Technology and obtained an Executive MBA from INSEAD.