Irrationality for the Semiconductor Industry

Michel DEPEYROT. CEO Dolphin Integration. September 2014. ShangHai.

Contemplating the last 50 years takes us back to California, more precisely to Stanford University and to the year 1964, three years after Carl G. Jung's death in Switzerland, founding father of the psychology of cultures and our best guide to the deep unconscious.

Over the sixties, pioneer Fairchild Instruments sired Semiconductor start-ups over "the Valley" around Stanford University, which staffed their research teams with engineers skilled in design and fabrication of MOS Silicon Gate transistors.

1964 is the year when my thesis advisor Michael Arbib published his book on "Brains, Machines and Mathematics". Neurosciences then were emerging at the same time as the psychology of Transaction Analysis and the Semiconductor Industry.

As we wish to believe our industry to be driven by rational moves, it is worth pondering some of the seeds of the sixties or soon after, which are worth tracking for their irrational impact on our semiconductor industry to this day.
Not that being rational would be equated with goodness, versus irrational with evil, but they rather are related respectively with being less risky, versus unsafe.

Let us start by distinguishing three kinds of irrationality:
• individual irrationality, which may be due to insufficient breath of knowledge, to the dominance of emotions, to some lack of conviction, etc.
• while an industry is concerned with collective or corporate irrationality, which may be caused by the fear of competition, by stampedes due to the need to track fads, or to irrelevant financial rules imposed by investment actors, etc.
• and partnerships are mired by relational irrationality, beginning with the cultural differences in negotiating styles, all too clear between Occident and Orient, etc.

There is much to gain from a better awareness freeing us from what may be blinding, or what may be more likely leading us to wrong moves.

Let us review a diversity of issues, where some encompass technological mastery, some have to do with marketing savvy, others with financial skills, and all of them require improvements in both reasoning and intuition.

#1. Irrationality in foresight of disruptive innovations

As early as 1964 the move to parallel computing was started with the confidence that computing had to imitate brainwork and that cellular automata would do the trick…

As for Threshold logic, it resulted from Cowan and Winograd's simplistic model of firing neurons: it did not go as far as it promised…
Then Neuronal networks in turn announced astounding processing capabilities, just to yield narrowly specialized algorithms…
Still today some advanced research again is aiming at designing an IC to emulate the way the brain supposedly works: they look at the brain hardware through MRI, but they ignore its actual workings, they ignore the way brain and body interact to trigger the workings of the imaginary, that brain software, as distant from the workings of a Harvard architecture in a data-processing machine as from the workings of Threshold logic or Neuronal networks.
The flaw in the analogy between Brains and data-processing Machines had been perceived as early as 1918 by Carl G. Jung, namely the omission of the fact that the brain-body system processes psychic energy, while computing processors only deal with data.
Indeed psychic energy - like any energy - is the product of an intensity and an extensity, e.g. an instinct and an emotion.

Donald B. Calne, wisely reminds us that "the essential difference between emotion and reason is that emotion leads to action, while reason leads to conclusions".
Meanwhile, between conclusions and action, there must happen a decision, and free-will must intervene.
The slant to Intuition – our most irrational source of knowledge - may guide some persons away from the opposite way of the thinking-oriented person, as these orientations appear on the quadrant of psychological types defined by Carl G. Jung.
The other two such types have the slant toward affects, thus are sensation-oriented versus value-oriented (so-called feeling-oriented), so that you can bet that
the "sensation-oriented" person will do nothing until having seen a prototype demo,
the "value-oriented" person either, until a proposed endeavor is proven to fit the chosen strategy.
As for the "thinking type", unless this person is presented with proper facts and with an RoI analysis, don't count on any action, while the "intuition-oriented" type shall refuse any move or else shall surprisingly go ahead, only based on some deep hunch.

Moreover it does not suffice to face or to share factual knowledge to reach a decision, decision criteria also must be faced or shared.
While a decision made on a basis of explicit criteria is seen as rational, it may still be missing that hidden quality, which would make it turn into action. A decision is seen as irrational simply when it is made on a basis of unconscious criteria; it suffices that some hidden objection ultimately gets expressed, and any action gets inhibited.

Back to the three types of situation:
• When individual decisions are concerned, intuition may yield no action or surprising actions. Some consulting help may come from using a decision matrix and a Figure of Merit with some strenuous effort to quantify intimate criteria.

• When group decisions are involved, as often needed within a corporation, peer pressure may lead to the fear of expressing objections, resulting in apparently rational but wrong decisions and action. Methodical setting of a "decision matrix" and assigning weights to its criteria may be done by voting, but the knowledge base of participants – when it is diverse enough – introduces a dangerous inconsistency between driving forces of the four types: thinking, feeling, perception and intuition.
Collective decisions are so unlikely: **visionaries must be lonely decision-makers!**

- When negotiations are involved between companies attempting to make a deal, new consulting skills are involved as success comes from identifying a win-win compromise respectful of the cultural slants. The need of an explicit and shared Figure of Merit is made tougher as each party may not want to openly share their hidden agenda.

#2. Irrationality in the forecast of technological shifts

Let us start with an easy one, the selection of material for the substrate of integrated circuits: the most rational choice in the sixties was not silicon, but germanium. Nonetheless silicon benefitted from the omitted cost advantage, which compensated all its disadvantages, while the underlying fabrication cost issue was to become the prevalent obsession of this semiconductor industry.

Let us go-on with a popular case, namely the edict of "Moore's law": **here is an edict which both is a personal proposal, and is based on intuition**, upon some brief observations as there was yet no history of the semiconductor industry. The claim was that the number of transistors in an IC would approximately double every two years. It was foreseen by a co-founder of Intel, Gordon E. Moore, whose description of the trend was published in 1965.
His prediction has been proven accurate or self-fulfilling, because every actor in the semiconductor industry knew to have to set targets per this demanding law, lest being overrun by the competition and by Intel to begin with.
This forecast exponential improvement has dramatically enhanced the impact of digital electronics in nearly every segment of the world economy, for it became forever the substitute of rational collective decisions of our industry.

But it does not work always as well! Fifteen years later came the shift from NMOS to HCMOS in 1980, already aimed at reducing power-consumption, while the decline of SOS triggered the pronouncement that SOI was going to be unavoidable soon, as it still is today, 35 years later, "soon unavoidable"…

It is interesting to observe that SOI was starting to be attractive at 28 nm just at the point where a drastic change had to happen, lest Moore's law would not pass the 10 nm barrier. What happened is that the last half-century of electronics (PCBs) and microelectronics (ICs) has grown on the basis of 2D printing.
What escape route was to be found for Moore's law to continue: here it is, the third dimension. This is what FinFET is announcing and a death knoll for SOI at once.
While the future is 3D printing, SOI is clinging to a 2D fabrication principle.
Meanwhile, we have to rethink ALL our design techniques, from memories to analog devices with a completely renewed paradigm. The first 3D analog amplifier is the next challenge.

Then let us now take the wicked case, namely the pronouncement that mixed signal designs would no longer be viable at the next fabrication node. I remember well how a semiconductor strategist in Japan, 20 years ago, at the time when analog virtual components were just starting to appear embedded in some mixed-signal circuits at 350 nm (.35 μm), was forecasting that it would be the last node in which such a mix could be performed profitably. Now we know that 28 nm processes are doing just that again…
What will rationally keep Moore's law and the More-than-Moore corollaries working in 2024 with the higher fabrication volumes demanded for 10 nm processes or beyond?

#3. Irrationality in the acceptance of unavoidable industry trends

As early as 1969 Jerry Sanders created AMD to compete in Intel's race for **Int**egrated **El**ectronics.
Then came another case of a truthful forecast: "Silicon compilers and foundries will usher in user-designed VLSI" article by Carver A. Mead and George Lewicki. Caltech in the "Electronics" magazine of 1982 Aug 11.
This article was announcing the unavoidable double impact of deverticalization of the semiconductor industry, rationally justified for both the design industry and the fabrication industry.
Carver Mead, who had coined the term "Moore's Law", had himself started a foundry service for researchers to share the cost of manufacturing prototype semiconductors, a program which inspired the industrial foundry services.
Now, deverticalization certainly was on its way, but still is far from completed 30 years later, after numerous investors have contributed to resisting it.

And the country who has most efficiently accepted it, is its late comer, China.

On their way to focus better on product innovation, some initially Integrated Device Makers (IDM) have progressively stopped investing in advanced fabrication processes and have turned Fab-Lite before the more radical move to becoming a pure-play Fabless supplier.
The second and simultaneous economic shift is adding efficiency, by focusing on short Time-to-Market through acquiring Virtual Components of silicon IP (ViC) and implementing a complex System-on-Chip (SoC) by embedding more and more productized silicon IP.
It took until 1984 for the emergence of Silicon Compilers as "Memory Generators" and of "Logic Synthesizers" for designs based on Register-Transfer Languages (RTL).

But a hidden economic shift was happening at the same time, due to the integration of system functions within ICs. Numerous Fabless suppliers and the growing complexity of circuits per Moore's law have resulted in replacing electronic cards of devices based on PCBs by SoCs based on silicon: actually the role of standard ICs to produce devices based on PCB's has been replaced by the role of standard ViCs in the fabrication of SoCs.
Surreptitiously came the time of reckoning: instead of buying and paying IC's progressively as the ramp-up of device production demanded it, Fabless suppliers – driven by American CFOs oblivious of economics – claimed to want to pay ViCs as Non-Recurring Expenses (NRE) only (no royalty compensation), causing a brutal cash problem for the semiconductor industry:
it was tantamount to wanting to put in stock all ICs in advance of volume production.
In contrast, China has been the leader in this quandary of royalties versus NRE, which globally paved the way to a more reasonable compromise for the cash spending pattern.
The irrational refusal of royalties by some financial "experts", versus their acceptance in China for reducing NRE, and not just the temporarily lower wages, explains to a large extent the shift of the vortex of this design industry across the Pacific Ocean.
At the same time, the emergence of Virtual Components of Silicon IP and the acceptance of such a scheme to enable Design for Reuse or D&R, contributed to the unique drive to costing-down, which characterizes the semiconductor industry.

This drive had been started in 1974 by Texas Instruments with its use of the Fabrication Learning Curve to forecast the advantage of quantities to drive IC costs down.

Similarly irrational down-sizing by innovative Redesign is more efficient and a better path to success than cutting margins… Indeed reducing fabrication cost or power-consumption is best conducted irrationally, as yield increase is more rewarding than area shrinking, etc.
A typical example for the worth of a search for cost-reduction based on innovation is given by software solution BassPower™ enabling the elimination of a big capacitance severely impacting the bill-of-material.

#4. Irrationality about intellectual property and innovation funding

IP-97 dates the official appearance of this strange IP acronym, at the first of a series of technical conferences, starring Robin Saxby who announced the emergence of a new industry driven by ARM Holdings, that of semiconductor design which moved from an era of pure service to an era of product offerings.
The professional association wisely named VSIA (Virtual Socket Interface Alliance) coined the term "Virtual Component" for truly productized IP.
Governments and many companies, whatever the industry concerned, are involved in a furious race for patenting, while there are two other less simplistic strategies for protecting innovation or Intellectual Property.
• The oldest defense is secrecy, which is applicable to fabrication techniques, which can only be countered by illegal spying, and for which the new business of circuit security has emerged.
• The most sportive defense is the rate of innovation itself, made feasible by Moore's law, which entails to unavoidably disclose innovative products and to courageously race to introduce further innovations. It also requires flawless execution.

The deterrent in the patent game… versus the push for patents, is the lawyers' way, at large portfolio owners with deep pockets, for heavy legal fees, of not respecting others' patents and suing the smaller actors on any fancy counter-accusation of patent violations.

Zhao TingYang: "Du ciel à la terre : la Chine et l'Occident"
"… I am afraid also that commercial protectionism, **the monopolistic rights on knowledge and technique**, the multiple taxes shall bear a part of responsibility in the increase of worldwide economic disorder."
Accelerating innovation, lest some competitor shall overcome you by not respecting your patents and IP anyhow, has been awesome in the country which still is the least settled, China.

There is more to it: venture-capitalists act in irrational denial of a basic law of capitalism.
Take a small enterprise short of capital but on the right market, still a niche unknown to most market analyst, with a vision and with inner capability:
financial advisors demand to prove first the existence of this market by waiting until competitors appear… But the entrepreneurial drive to innovate and grow on these premises entices this corporate leader to invest heavily for mid-term return – say two-years – at the risk of not yielding profits during this pre-growth phase. It should be the right time for an increase of capital, but the valuation is abysmal, as the books show no difference with a company on a hopeless track.
The likelihood is to loose corporate control to the benefit of market-blind financiers. The entrepreneur's only escape is to manage for survival through these hard times with cash-crunch, and if so, investors shall then want to invest at no risk but to lessen the return on equities… This is the history of historically best innovations, on a shoe-string.

Let us look at the seed funding of the Semiconductor top-3 contenders today for the leadership in microprocessors for tablets:
1968. Intel has to invent Stock-Options.
1984. Qualcomm is started, as service supplier, by seven experienced professionals with own funds.
1990. ARM Holdings is funded by Apple Inc. with their own capital.
Once big, they keep their rank with the strategy of domination, keeping their advance from the start by their rate of innovation.

The only result European so-called Venture Capital is good at is selling European jewel start-ups to American companies. And they influence so well the administration that they get delegation for such misuse of national funds.
They refuse precisely niches "they had never heard about" (quote) or "they do not understand" (quote), like RISC machines in 1984 or the IP business in 2014. At least, they are predictable!

#5. Irrationality in applying complex mathematical theories
The year 1964 was also the time of booming for a theoretical breakthrough: Shannon theorem and Information theory for communications. It led to the claim that modems could not reach transmission rates above 30 kiloBauds.
Nonetheless for a given modulation rate (in bauds, to quantify how fast the signal value changes), the data rate can be increased by changing the encoding method. The maximum channel capacity could actually go far beyond 32 kb/s thanks to phase-modulation… Just for having confused Bauds and bits/second in the theorem…
Today the maximum Baud rate currently achievable over fiber cables is 3 Million Bauds.
Meanwhile huge investments have been made for the development a fast digital alternative: **Integrated Services Digital Network**. ISDN is basically a set of communications standards, which allow the transmission of network services, including data, video and voice. It is a digital standard which uses a dedicated link for each exchange. It can typically provide around 128 kbits/sec. of simultaneous upstream and downstream throughput, certainly far better than the early modems could do on telephone lines, but it has nowhere near the capabilities that **Asymmetric digital subscriber line** does. Instead of relying on a special carrier, back to copper telephone lines, ADSL is a data communications technology, which enables faster data transmission than a conventional modem over the voice band, can provide, simply by utilizing frequencies which are not used by voice through a telephone line.

The semiconductor industry happily provided all the integrated circuits for all these generations of communication techniques, like ISDN then ADSL, up to making the Internet viable at low cost.
Indeed, the traditional telephone network, which ISDN was developed around, is circuit switched. When you call someone on a circuit switched network, a set of wires is connected between your telephone and the callee's telephone. By contrast, the Internet is a packet switched network. Instead of creating actual wired connections between two terminals, all locations are interconnected all the time. The information on the packet switched network is in the form of individually addressed packets.

As the time came for machine-learning, so came the new human-inspired solution of **Artificial Neural Networks**. ANNs are computational models, supposedly inspired by the brain works, to give the capability of pattern recognition to a robot. The ANN is generally presented as a system of interconnected neurons, which can compute values from their inputs.

But the brain has two major differences with such a device: on the one hand it is a dual network of gray cells (neurons) and white cells (astrocytes), where the latter drive the permanent restructuring of the gray cells, up to the plasticity of their genes, and on the other hand each of the neuron interconnections involves chemical synapses, where a flow of neurotransmitters is brought by astrocytes to control the transmission patterns. So that the "firing" must not be over-simplified, as if it were about an electronic network.

#6. Irrationality in the fate of computer architectures

Here the challenge is about Standard products versus Custom design, which ultimately turns into cost minimization versus performance maximization.
A company created in 1968, namely Intel again, came-up with the first standard single chip microprocessor per the Harvard architecture, the 4004 in 1971, which less than ten years later sired the 8080 and the 8051. It was then called **MCS-51**, with a complex instruction set (CISC), a single chip microcontroller embedding a configuration of peripherals for general-purpose control. Intel's original releases were popular in the 1980s and early 1990s and have largely been superseded by a vast range of faster and/or functionally enhanced 8051-compatible devices, which have been fabricated by more than twenty independent Integrated Device Makers (IDM).
But a new life has been given to this architecture with the advent of the IP business in the late nineties, to the point that its latest successor today relies on the original CISC structure plus some blending of its instruction set with best concepts from RISC structures.
The major competitor came from Motorola as the 6800 microcontroller far ahead of the 8051, in 1974, as its architecture and its instruction set were inspired by then so popular PDP-11 mini computer of Digital Equipment Corporation (DEC). The 6800 derivative MC6805, embedding RAM, ROM and I/O on a single chip, was successful in automotive applications and communication display terminals:

- Who was going to be the winner, and why?

The 8051 is still alive and kicking because it was neither a Customer Specified Integrated Circuit (ASIC), nor an Application Specific Standard Product (ASSP), but a standard circuit for which a flurry of software developers created an immense data base of programs, ready for reuse. It also became the standard-bearer for the SmartCard business worldwide.

To no avail: the **A**rgonaut **R**ISC **C**ore (ARC) tried the same differentiation against the ARM (Acorn RISC Machines). Their core was created in such a way that it is quite rationally extensible. **ARC International** was a developer of configurable microprocessor technology with synthesizable IP and licensed it to semiconductor companies able to perform some design, while ARM was licensing its core to foundries unable to do so.
Same crucible for Digital Signal Processors (DSP), modifiable versus standard.

Configurations of the ARC happen at design time (as opposed to run time). Unlike most embedded microprocessors, extra instructions, registers and functionality can be easily added. In view of the task to perform, users assemble their own custom microprocessor, optimized for speed, energy efficiency or code density. But they end-up with an ASIC, with an ASSP at best, not with the proliferation potential of a standard microprocessing product, due to the incompatibilities in diverse software developments.
As the cost of software now is higher than the cost of semiconductor hardware, the crucial challenge is to innovate, while preserving at best the software compatibility over product generations: a startling case for the merger of opposites…

On the basis of the more than 12 Billion ICs fabricated for the 80C51 and its legacy products as ICs, not counting the SmartCards, the MCS-51 and its CISC architecture of 1980 with the addition of some RISC flavor, now is siring ZEPHYR, which is daring the challenge for the era of the Internet of Things, together with SmartVision, a block-busting platform for software development. Long life to it!

#7. Irrationality in the opposition between completeness and efficiency

Rational Occidental thinking is based on binary logic at the same time as philosophical reasoning is based on the triptych of "Thesis-Antithesis-Synthesis", but it took until Dutch mathematician Luitzen Egbertus Jan Brouwer to challenge this domination of Cartesian thinking.
Brouwer lived from 1881 to 1966. He was the founder of the mathematical philosophy of intuitionism, which is sometimes characterized by the refusal to use the principle of the "excluded third" in mathematical reasoning. In a 1908 paper entitled "The untrustworthiness of the principles of logic", he challenged the belief that the rules of classical logic, which have come down to us essentially from Aristotle's principle of non-contradiction (384--322 B.C.), through René Descartes, have an exclusive validity. He highlighted the dependency on the subject matter to which the application of binary logic is attempted.
The search for the hidden third and its inclusion has been promoted by Franco-Rumanian Stéphane Lupasco: the drive is about getting the best from the harmonization of opposites, which yields a new value, which in turn shall go to an extreme, so as to trigger the emergence of its opposite…
Let us see how the opposition between completeness and efficiency yields Simplicity.

The tenants of complete architectures in the late seventies pushed their designs beyond the celebrated Motorola 68000, endowed with the most CISC architecture (Complex Instruction Set Computer) and a number of opposites were launched by Berkeley's David Patterson (who coined the term RISC for Reduced Instruction Set Computer) between 1980 and 1984 in their search for efficiency. The trade-off is about compiler efficiency involving the speed of instruction decoding versus the speed of operations. The alternatives can be compared on histograms for the actual rate of use of each instruction in some application.

2004. Similarly the Swiss research center CSEM launched in the new century the concept of a Reduced Cell Stem Library (RCSL) versus the previous century's trend toward ever more complete CCSL (Complex Cell Set Libraries). The trade-off is about synthesizer efficiency, involving the feasibility of selecting the optimal cell for placement, versus the spacing required by routing. The alternatives can be compared on histograms of the actual rate of use of each cell in some SoC, depending on the optimization in speed or density.
The opposition between complex and reduced design solutions invites to uniting "diversity of optimizations" with "simplicity of design by assembly".

2014. The next frontier has to do with energy preservation, but then it will take a wise combination of software, analog design and logic design for moving away from any current CPKL, both Complex and Custom Power Kit Libraries to a block-busting RPKL, both Reduced and Reusable Power Kit Library, similarly opposed to « endless incremental complexity ». The search for the optimal Power Management Network of some SoC mixes a number of criteria: efficiency, density, leakage, dynamic power, noise, bill-of-material… obscuring comparisons.

So what?
Beware the stampede of irrationality triggered by Marketing for capital investors.

Let me now give the beginning of **Zhao TingYang's quote:**
"For sure, the today's worldwide economic crisis is first due to the movements of financial capital swollen and beyond control, but I am afraid…

• the PC era yielded productivity increases by information processing in the seventies and eighties;
• the Telcom era yielded productivity increases by Internet enterprises in the nineties and the 2000 decennia;
• the IoT announces information systems in smithereens – be it for obsessional tracking of health, for wearable snobbism, or for distributed control – certainly with high fabrication volumes and thus high potential, but without any clear focus on productivity increases:

- Who is going to bring the worthwhile innovation for the next twenty years?
- What shall be worthwhile for the world economy?

We may at this stage foresee only two potentially significant contributions to worldwide productivity and even survival, namely radical energy saving. But we are already confronted to opposites:

→ where SmartGrid only aims at eliminating the peaks of energy, and thus the energy providers' investments
→ while SmartHome only aims at reducing the energy spending and improving comfort.

At least they will both contribute to pollution control.

Let us optimistically count on further irrationality of our industry to do what is right, rather than on decisions by rational investors. Indeed, "the deficiencies of thought are clogged by appropriate actions" [Zhao TingYang].