CSE 40547/60547 – Computing at the Nanoscale– Spring 2011
Homework 01 – Practical Limits to CMOS Scaling
Assigned: January 26, 2011 **– Due:** February 2, 2011 (by email)

During the second and third weeks of class, we have been discussing limits and issues associated with CMOS scaling. The main purpose of the lectures has been to introduce the fundamental parameters associated with several different scaling models. While the models give us a good understanding of how smaller device dimensions and changes to the supply voltage will ultimately affect performance (by which I mean both latency *and* power) much of the discussions have ended with first level approximations – necessary to make the material associated with this very complex topic fit into just 5 lectures!

In order to gain a better understanding of more specific trends – and how scaling will affect performance at the application level – in this assignment you will listen to a lecture by David Frank of IBM (a short biography is include below).

Dr. Frank received his B.S. from the California Institute of Technology, Pasadena, CA in 1977 and a Ph.D. in physics from Harvard University, Cambridge, MA in 1983. Since graduation he has been employed at the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he is a Research Staff Member. His studies have included non-equilibrium superconductivity, III-V devices, and exploring the limits of scaling of silicon technology. His recent work includes the modeling of innovative Si devices, analysis of CMOS scaling issues such as power consumption, discrete dopant effects and short-channel effects associated with high-k gate insulators, exploring various nanotechnologies, investigating the usefulness of energy-recovering CMOS logic and reversible computing concepts, and low power circuit design. Dr. Frank is an IEEE Fellow and has served as chairman of the Si Nanoelectronics Workshop and is an associate editor of IEEE Transactions on Nanotechnology. He has authored or co-authored 100 technical publications and holds 10 U.S. Patents.

More specifically, David talks about "The Limits of CMOS Scaling from a Power Constrained Technology Optimization Perspective." In less complex terms, this talk begins by essentially considering most – if not all – of the same parameters that influence performance that we have considered in class. Using more detailed models, he comments on ideal feature sizes, to obtain maximum performance, for different classes of applications.

For this assignment, you'll need to listen to the lecture and answer the questions on the second page of this handout.

You can find an .mp4 file of the audio + slide transitions linked off of the course website.

You can also find a PDF of the slides (that should make following along with the presentation quite a bit easier) linked off of the course website.

Some questions can be answered with just a few sentences. Others require about ½ a page to realistically be thorough. Your answers do not need to come solely from Frank's presentation. Referencing material covered in lecture, industry roadmaps, etc. is also a good thing to do. (i.e. the ITRS executive summary is an excellent resource for answers to a few questions!)

Question 1:
What thermodynamic issues should we be concerned with as feature sizes scale?

Question 2:
The International Technology Roadmap for Semiconductors (or ITRS) is a set of documents produced by a group of semiconductor industry experts. These experts are representative of the sponsoring organisations which include the Semiconductor Industry Associations of the US, Europe, Japan, Korea and Taiwan.

The ITRS represents the best opinion as to the directions of research into the following areas of technology, including time-lines up to about 15 years into the future: (a) System Drivers/Design, (b) Test & Test Equipment, (c) Front End Processes, (d) Photolithography, (e) IC Interconnects, (d) Factory Integration, (e) Assembly & Packaging, (f) Environment, Safety & Health, (g) Yield Enhancement (h) Metrology, and (i) Modeling & Simulation.

You can find the Executive Summary of the ITRS roadmap linked off of the course website or at:
- http://www.itrs.net

In the context of the scaling models discussed in class, realistic power budgets, etc. comment on chip power density if the trends in the roadmap play out as predicted.

Question 3:
What determines V_t (the threshold voltage)? Also, comment on the importance of, and methods for, scaling V_t .

Question 4:
According to David Frank, what is the most important optimization constraint when considering device dimension scaling? Why?

Question 5:
Frank talks about concerns over increasing leakage currents and increased power as a result. Why is increased power from leakage currents, relatively speaking, bad? (Hint – yes, temperature will increase but this is not the answer that I'm really looking for.)

Question 6:
Frank talks about increasing performance by increasing mobility. Discuss one way that we can increase mobility. (Suggestion – don't just write down a term that's used in the talk. Do just a bit of independent research to answer this question more thoroughly. Some useful papers have been linked off of the course website.)

Question 7:
How will performance scale as power increases? How has this affected chip architectures in recent years?