GPU Accelerated SPICE Simulation for Analog IC Designs

Empyrean Software

SPICE (Simulation Program with Integrated Circuit Emphasis) has become an indispensible tool for the simulation of transistor-level circuits since its introduction in the early 1970s. Over the years, many SPICE simulators have been introduced and their capabilities have been greatly improved. However, as we move into deeper sub-micron designs and circuit sizes keep increasing, the capabilities of current SPICE simulators prove insufficient; continuous performance breakthroughs are needed to keep up with the demands of larger circuits, high accuracy and fast turn-around.

Given a netlist that describes the topology of a circuit, a SPICE simulation job consists of several major tasks. Netlist parsing and building database are usually done sequentially. Then the simulator can use parallel processing techniques, such as multi-core and multi-threading, to speed up tasks for calculating DC operating point and transient analysis. Model evaluation is another task that can be easily parallelized, although it is usually a small percentage of the total simulation time. All modern SPICE simulators use parallel processing for speedup, but they differ in the techniques used and can lead to very different capacity and performance.

Empyrean ALPS™, which stands for Accurate Large-capacity Parallel Spice, is a parallel SPICE simulator that employs advanced partitioning and breakthrough techniques for matrix solving, and offers the best capacity and performance compared with other SPICE simulators in the market.

Empyrean ALPS™

The major techniques used by Empyrean ALPS™ are explained in the following. For a given circuit, ALPS first partitions it into overlapping blocks using a hyper graph partition method. Then, BDF (Backward Differentiation Formula) methods are used to discretize the set of differential algebraic equations. For each block, the internal matrix is usually extremely sparse, where LU-decomposition is a very efficient method to solve it, and it is also suitable for applying parallelization techniques.

The matrix for the block coupling nodes is usually dense, for which GMRES is the most efficient method. However, a major flaw of the GMRES method is that it needs an efficient pre-condition matrix; otherwise it will converge very slowly or cannot converge at all. It is not easy to select a good pre-condition matrix, and even with a good pre-condition matrix, it often takes many iterations to solve the equations successfully. A breakthrough in ALPS is that we have developed new techniques that can not only get an efficient pre-condition matrix, but also enable GMRES to converge quickly in most cases -- therefore greatly reduce the time to solve the equations.

Algorithm: GMRES – with (right) Preconditioning

1. Initialize x_0 and a dimension m of the Krylov Subspaces.

2. Execute Arnoldi process:

   - Set $r_0 = b - Ax_0$, $\beta = \| r_0 \|_2$ and $v_1 = r_0/\beta$
   - For each $j = 1, \dots, m$
     - Compute $z_j := M^{-1} v_j$
     - Compute $w := Az_j$
     - For $i = 1, \dots, j$, do : $\begin{Bmatrix} h_{i,j} := (w, v_i) \\ w := w - h_{i,j} v_i \end{Bmatrix}$
     - $h_{j+1,1} = \| w \|_2$; $v_{j+1} = w/h_{j+1,1}$

   Set $v_m := [v_1, \dots, v_m]$ and $\bar{H}_m = \{h_{i,j}\}$

3. Set $x_m = x_0 + M^{-1} V_m y_m$

   Where $y_m = argmin_y \| \beta e_1 - \bar{H}_m y \|_2$ and $e_1 = [1, 0, \dots, 0]^T$

4. If satisfied stop, else set $x_0 \leftarrow x_m$ and goto 2.

The capacity and performance of ALPS make it suitable for simulating large post-layout circuits. For such circuits, ALPS typically gets 2-3X speedup compared with other true SPICE simulators. So far ALPS has hundreds of paid customers worldwide.

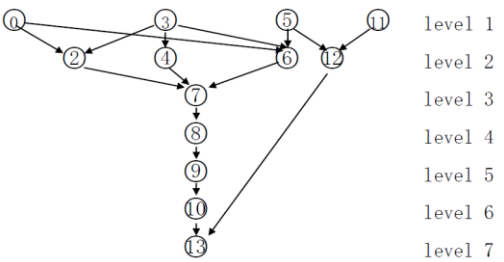
Empyrean ALPS-GT™

Besides using multiple CPU cores to speedup circuit simulation, using GPUs is another way to further speedup simulation by hardware acceleration. For tasks such as matrix solving, which is usually the bottleneck in the whole simulation process, if we can do it effectively using multiple GPUs, very significant speedup can be obtained.

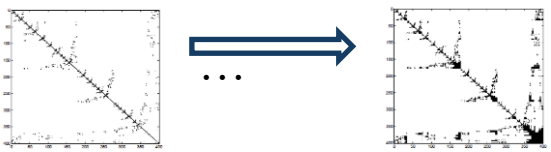
However, matrix solving on GPUs is a challenging task and many SPICE simulators have tried and failed. Taking the Gilbert-Peierls left-looking algorithm as an example, it needs to scale sparse vectors to dense vectors, thus a large cache is needed. The LU-decomposition solver also requires large cache. But usually cache size on a GPU is much smaller than that on a CPU, so we need special techniques to perform these tasks on GPUs without being limited by the cache size and data transfer overhead.

Empyrean Software has developed many innovative techniques to speedup ALPS by using multiple GPUs, and the company released Empyrean ALPS-GT™ in the 2019 DAC. Illustrated in the following graph is one of the techniques that combines left-looking and right-looking algorithms for faster matrix solving.

Lower triangular substitution jobs generation by dependence tree from topological sorting of U matrix
- Jobs of next level have to wait until jobs of previous level are done
- Both Level 1 and level 2 have four parallel jobs
- A large number of left looking jobs can be generated from extremely sparse matrix
- Use left looking method to handle huge number of parallel jobs



Level 3 to level 7 have only one executable job
- Matrix will be relatively dense with mode nodes being eliminated
- Most left looking jobs will be executed in serial due the dependence tree
- Use right looking method to generated more parallel executable jobs for nearly dense matrix



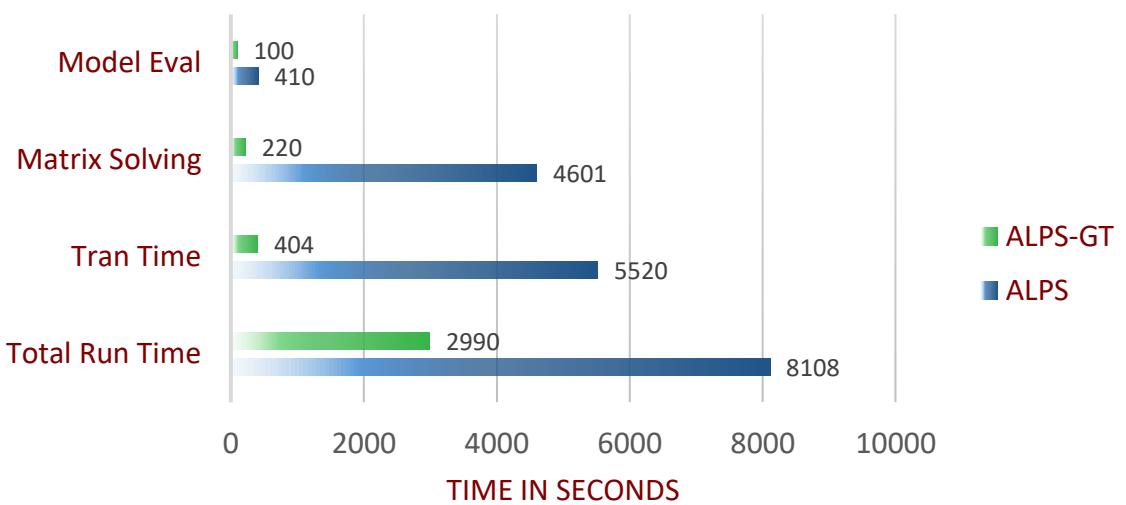
Benchmark

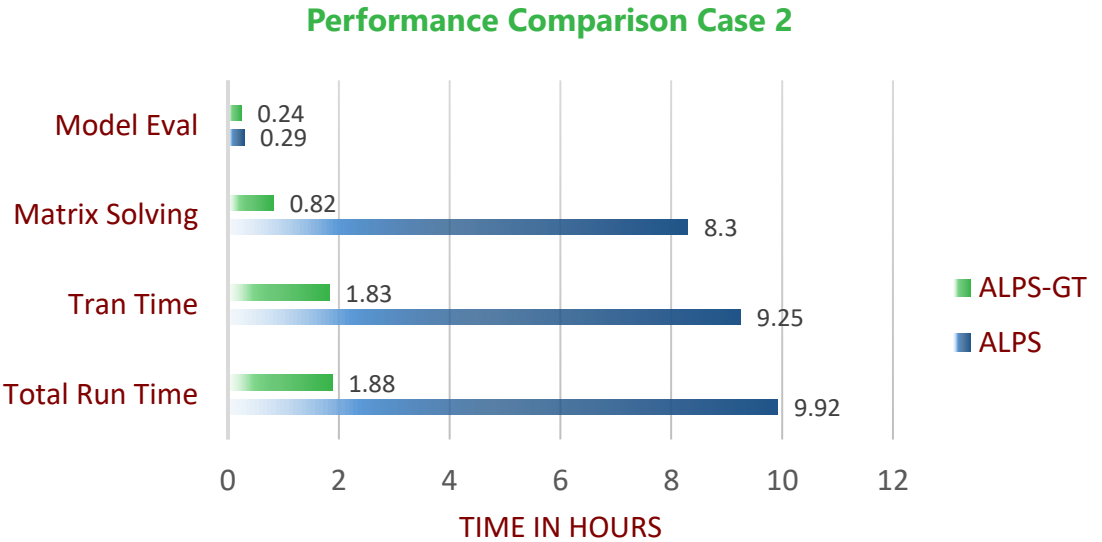
Empyrean ALPS-GT™ has been adopted by leading SoC design companies and has shown significant speedup especially for deep sub-micron analog designs in 16/7/5nm process technologies. In a recent comparison done by Nvidia' s engineering team[1], an Nvidia DGX-1 server with 2 * Intel Xeon E5-2698 V4 20-Core CPU and 8 * Nvidia Tesla V100 GPU is being used.

Simulator	CPU	GPU	Acceleration Configuration
Empyrean ALPS™	Xeon E5-2698 V4 Dual 20-core Intel Xeon		8 CPUs
Empyrean ALPS-GT™	Xeon E5-2698 V4 Dual 20-core Intel Xeon	Nvidia Tesla V100	8 CPUs + 8 GPUs

Several post-layout design blocks are simulated using both ALPS and ALPS-GT on the hardware platforms. Shown below are the results for two blocks that are representative of the overall comparison.

Performance Comparison Case 1





Conclusion

Empyrean ALPS-GT™ is the first commercial GPU-powered SPICE simulator. It can greatly improve simulation performance when the time for transient analysis and matrix solving dominate total run time, which is usually the case for many large and post-layout circuits. It is easy and seamless to use Empyrean ALPS-GT™ in an existing design environment, and benchmark result shows that Empyrean ALPS-GT™ can deliver 5~13X speedup over CPU-based SPICE simulators without any accuracy loss.

References

[1] Chen Zhao, "GPU-Accelerated SPICE for Deep Submicron Analog Simulation", GPU Technology Conference 2020. Available: https://www.nvidia.com/en-us/gtc/posters/?search=P22496

Sales Contacts

Headquarters
Empyrean Software

2F Building A, Wang Jing Hi-Tech Park,
No.2 Lizezhonger Road, Chaoyang
District, Beijing, 100102, P. R. China
TEL: +86-10-84776888
FAX: +86-10-84776889
Web: www.empyrean-tech.com
EMAIL: info@empyrean-tech.com

USA
Empyrean Software
4030 Moorpark Ave, Suite 109, San Jose,
CA 95117
TEL: +1 (669) 230-5019
www.empyrean-tech.com
E-mail: info@empyrean-tech.com

Korea
linkGlobal21
#301, 81, Hyeonam-ro, Suji-gu,
Yongin-si, Gyeonggi-do,Korea
Tel: +82-70-5138-0700
www.lg21.net/
E-mail: eda@linkglobal21.com

Singapore
MEDs Technologies Pte.Ltd.
5012, Ang Mo Kio Avenue 5 #04-01
Techplace II, Singapore 569876
Tel: +65 6453 8313
Fax: +65 6453 7738
www.meds-tech.com

Japan
Xloud Inc.
Dojima Axis Building 4F,
2-2-28 Dojimahama, Kita-ku,
Osaka, 530-0004 Japan
Tel: +81-6-6123-7784
https://xloud.co.jp
E-mail: eda_sales@Xloud.co.jp