

Large Scale NVH analyses for General Motors Using Cray SV1

Dr. Kristyn J. Maschhoff , Mr. Nathan L. Wichmann,
Dr. Himanshu Misra, Cray Inc., USA
Dr. Wayne Nack, General Motors Corporation

ABSTRACT

MSC.Nastran V70.7 was used on the Cray SV1 to perform large-scale NVH problems typically encountered in the automotive industry. Several MSC.Nastran solution sequences such as SOL 103, 111, 108, and 107 are heavily used by automotive customers in a production environment on the Cray SV1.

Significant performance improvements were recently made to key MSC.Nastran kernels to exploit the architecture of the Cray SV1 machines. Special attention was given to the normal modes (eigenanalysis) and frequency response modules invoked in typical NVH analysis. For example, proper blocking in the matrix decomposition and matrix multiply kernels allows most problems to achieve 75% of peak within these kernels by taking advantage of data reuse from cache. Suitable blocking allows maximal data reuse from cache and avoids bandwidth limitations. In addition, optimal settings of numerous MSC.Nastran parameters were established for a wide variety of NVH problems.

Major improvements in performance were demonstrated for a collection of real life large scale problems provided by General Motors. Results clearly indicate that the Cray SV1 is a robust cost effective NVH engine well suited for a production environment in the automotive industry.

Automotive NVH Analyses

Noise, Vibration and Harshness (NVH) analysis is a key automotive industry application. Full vehicle simulation, acoustics and frequency response calculations are routinely performed in order to design vehicles with improved ride and handling characteristics. NVH analysis is used to predict the response of a structure to imposed excitations or loads given appropriate boundary conditions. The sources of dynamic excitation can be external or internal to the vehicle [1]. External forces include road-induced shake/noise and aerodynamic effects. Internal forces include power-train reaction forces, tire/wheel imbalance and brake-induced forces. Using full vehicle, low frequency NVH models, these forces are used to predict and assess, for example, a vehicle's shake and boom response at various locations in the vehicle.

Typical NVH problems are computationally intensive because of the size of the models and the wide range of excitation frequencies required. For NVH optimization, forced response analysis and sensitivity calculations must be performed at hundreds of excitation frequencies. In a detailed full vehicle simulation, problem sizes on the order of 2-3 million degrees of freedom are used in practice. If modal frequency response is used, the number of eigenvalues (modes) extracted can regularly exceed 1000. Such large I/O intensive jobs require powerful supercomputers with sufficient memory and fast I/O.

Cray SV1 computers are extensively used in the automotive industry for large NVH simulations using MSC.Nastran. Dramatic improvements in overall performance due to recent architecture-specific tunings of key components have greatly reduced turnaround time for large-scale NVH computations, enabling automotive engineers to reduce the length of the product design cycle.

Industry Background

Brake analysis and NVH optimization are heavily used in the design process at GM. Full vehicle simulation, acoustics and frequency response calculations are routinely performed in order to design vehicles with improved ride and handling characteristics. For full vehicle response problems, each component is divided into external superelements of size less than 1 M degrees of freedom. These externals are then assembled into a full vehicle system model. Various load conditions are then applied including frequency response and random road response to evaluate system design. Each superelement can require as much as 80 Gbytes of scratch disk and 0.6 Gbytes of system memory.

It is desired to raise the frequency content of all noise and vibrations models to the mid frequency range. On a body this range is 200 Hz to 800 Hz. This could require MSC.Nastran superelements to contain over 3 M degrees of freedom. Both modal and direct solutions may be used and there is a tradeoff in efficiency for large models. New methods to solve large linear system of equations are being developed. One class called domain decomposition dissects the equations into small pieces and each piece is reduced to the boundary. The combined residual is then solved. The method is very efficient for small pieces of the order of 1000 DOF each. P elements could be used to raise the frequency content of an existing mesh higher. For production models, p elements require more disk than is available on current GM computers. An alternative would use spectral elements to model this frequency range. A consortium by the Navy is examining the possible use of spectral elements. The results of this consortium have not been made public.

Statistical energy analysis techniques are used for high frequency problems. These mix analytical plate solutions, empirical and historical data, literature solutions and experimental to derive a semi analytical model for high frequency vibration. Although the high frequency models produce useful results, the basic formulation can not model mid frequency response. This is due to the fact that SEA contains only local modes in the formulation. An effort is underway to extend SEA into the mid frequency region by combining it with a FEA model. It would require a detailed FEA model for that purpose.

To account for a wide variety of manufacturing variations, a stochastic design procedure is needed. Stochastic design requires substantial computational resources as many independent problems (design parameters are varied randomly) are run simultaneously. Current trends toward multidisciplinary optimization further stress available computational resources. To achieve efficient turnaround for today's design process, computational resources must be able to both address the capability to solve very large problems and the capacity to run several of these large jobs simultaneously.

MSC.Nastran Optimizations on Cray SV1

The SV1 is the first in a series of scalable vector supercomputers developed by Cray. A fourth-generation CMOS vector system, the Cray SV1 is designed to handle a broad range of vector applications. SV1 processors are configured in a symmetric multiprocessing architecture similar to that used in the Cray T90 and Cray J90 series supercomputers and are scalable up to 32 nodes. This scalable vector capability means most jobs can be run on a single Cray SV1 node, providing maximum ease of use. A large Cray SV1 configuration can run multiple vector applications simultaneously, providing a cost-effective throughput solution. Very large jobs can be run across multiple nodes using either shared memory or message-passing programming models.



For many years, Cray parallel vector systems were designed under the philosophy that there is no substitute for bandwidth to main memory. However, the cost of providing high bandwidth has decreased more slowly than the cost of providing central processor capability. As a result, observed latencies to main memory have become more significant. For more than two decades,

scalar system architectures have used data caches to cope with this trend and provide the bandwidth amplification necessary for cost-effective high-performance systems.

In non-vector system designs, a data cache must be able to exploit both spatial locality and temporal locality. Spatial locality refers to the proximity of memory in the address space and temporal locality refers to the degree of data re-use. In non-vector systems, spatial locality is addressed by way of cache-lines, with 128 bytes being a typical width at present. Any cache miss will bring 16 64-bit words from memory to the data cache whether they are needed by the processor or not. Note that wide cache lines are not necessary to exploit temporal locality.

The CRAY SV1 cache is 256 Kbytes (32768 64-bit words) with a four-way set associative design. Thus, any memory address can be stored in any one of four different cache locations (sets). Spatial locality is generally expressed in a user's code as an inner loop (looping through sequential addresses) and, in most cases, this spatial locality can be exploited with vector loads. For this reason, the cache-line size for the CRAY SV1 is only 8 bytes, or a single 64-bit word. Consequently, there is no wasted cache-line bandwidth penalty for irregular strides, or for gather operations. For scalar loads, the CRAY SV1 has a 64-byte cache-line width designed to exploit any spatial locality in non-vectorizable constructs.

MSC.Nastran provides options to enable tuning specific kernels for cache-based architectures by selecting blocking methods and sizes. For many key kernels, the single parameter that determines data re-use is the Lanczos Recursion block size MAXSET, used in the Block Shift and Invert Lanczos algorithm implemented by the READ (Real Eigenvalue Analysis DMAP) module. Kernels affected by MAXSET include the forward and backward solve kernels (FBS), the orthogonalization kernels, and the AXPR kernels. On the SV1, the default block size in MSC.Nastran V70.7 is 9.

For each of these kernels the primary objective was to reduce required memory bandwidth. This has two benefits: first, lower bandwidth allows one to take advantage of the faster CPU of the SV1; second, system throughput is enhanced because there is less pressure on system memory. Though the optimal choice of block size is a complex problem, generally MAXSET values of 9, 12, and 15 work well on the SV1. Values greater than 15 provide little additional benefit in reducing memory bandwidth requirements and may adversely affect overall MSC.Nastran performance.

The biggest performance improvements were made in the decomposition kernels used in solution sequences 103, 111, and 108. In versions of MSC.Nastran prior to 70.7, a rank update of 2 was used. This meant that for each trip through the kernel, 3 loads and 1 store were required for every four flops. Two of the loads were for the vectors being multiplied, while the other load and one store were for the accumulation vector. For 70.7, the default rank update was changed to 16. The new kernel now performs 17 loads and 1 store in its innermost loop. However, 16 of these loads hit in cache, resulting in the memory subsystem seeing only one load and one store for every 32 flops, as opposed to 3 loads and 1 store for 4 flops on a CRAY-T90. The new kernel has one-sixteenth the memory bandwidth requirements and can achieve a high percentage of peak performance.

The MPYAD kernel Method 1 Storage 2 has also been optimized for the SV1. In the MPYAD module a number of long vectors are multiplied together for each pass. These vectors can be blocked in to smaller sizes for re-use. While the amount of data re-use is not large, the benefits are still appreciable on the SV1. Further changes were made in MSC.Nastran V70.7.3 to enhance shared memory parallel performance for this method.

Results

Results are presented for three different models provided by GM. These examples cover solution sequences 103, 108, and 107. These examples are representative of the current NVH work load at GM and demonstrate the suitability of the SV1 for NVH analysis in a production environment. Results presented were run on a 32 processor SV1 at Cray Inc. in Eagan, Minnesota and are undedicated timings.

The first example is a normal modes analysis problem. It is a SUV body plus frame and normal modes are extracted by the Lanczos procedure. The goal is to, subsequently, form a reduced CMS superelement for use in a full vehicle system level run. This model contains 1.7 million degrees of freedom and all frequencies < 400 Hz are requested. For this model SOL 103 is performed, and 2755 modes are extracted. Total CPU time is dominated by the READ module. Parallel results were obtained using a DMAP based frequency segmented parallel approach. The serial SOL 103 job was run using a single CPU using 200 Mw for Nastran memory. An additional 200 Mw was used for FFIO. Because of the large number of modes extracted this particular run is very I/O intensive. The serial run required 300 GB of disk and the amount of I/O transferred was 5,816 GB. The parallel SOL 103 job was run using 6 slave processes each using 2 CPUs. In general, the frequency segmentation approach used for the 103 parallelization is most effective when one needs to extract a large number of modes. The number of modes found per slave was between 448 and 467 and the extraction times for each of the slave processes was well balanced. The frequency segmented parallelization is limited to the READ module. Results for the normal modes run are shown in Table 1. Elapsed time refers to elapsed wall clock time.

Table 1: SOL 103, Normal Modes < 400 Hz

	SINGLE PROCESSOR	FREQUENCY SEGMENTED PARALLEL
Total Elapsed Time	31.3 hrs	14.3 hrs
Total User CPU	29.4 hrs	37.2 hrs
I/O Wait	1.4 hrs	3.4 hrs
READ -Elapsed	22.0 hrs	4.2 hrs

The second example is a large direct frequency response calculation. It is a full vehicle system model subjected to a harmonic tire unbalance. When the direct formulation is used the effect of all modes is present and residual vectors aren't needed. For large models, a direct solution is more efficient than a modal one. This model contains 946,453 degrees of freedom and 60 frequency steps are requested. Results are presented for a serial run using only kernel level parallelization (SMP) and a frequency segmented parallel run. For both the serial and the parallel runs kernel level parallelism was set to 2 CPU's. The serial run used 150 Mw for Nastran memory, 150 Mw for FFIO and required 12 GB of disk. The parallel SOL 108 job was run using 5 slave processes each using 2 CPUs. Each slave process used 75 Mw for Nastran memory and 75 Mw of FFIO. The DMAP based parallel implementation for SOL 108 is limited to the FRRD1 module. Total disk requirements for the parallel run increased from 12 GB to 44 GB.

Table 2: SOL 108, Direct Frequency Response

	SMP (AUTOTASKING USING 2 CPU'S)	DMAP LEVEL PARALLEL IMPLEMENTATION
Total Elapsed Time	13.5 hrs	4.3 hrs
Total User CPU	18.5 hrs	20.0 hrs
I/O Wait	0.4 hrs	0.9 hrs
FRRD1 -Elapsed	12.0 hrs	3.0 hrs

The third example is from brake squeal analysis. Brake squeal occurs when the frequency of the friction-induced vibrations falls in the 2-12 KHz range. Dynamic instabilities have been identified as the cause of the vibration and noise produced during brake squeal. The complex modes solution SOL 107 is performed to assess dynamic stability. This problem, provided by GM, performs brake squeal simulations up to 12,000 Hz. It is built using solid elements and contains approximately 120,000 degrees of freedom. The model consists of a rotor, pad, caliper, and mounting bracket.

Brake squeal analysis exercises the CEAD (Complex Eigenvalue Analysis DMAP) module in MSC.Nastran. Modes are found using a block bi-orthogonal complex Lanczos method. If there are no roots with unstable damping, the design is complete. For this input deck 110 modes were extracted in the range up to 12 KHz. Preliminary performance results are shown in table 3. Additional tuning efforts to the CEAD module are currently in progress.

Table 3: SOL 107, Brake Squeal Analysis, 110 modes

User CPU (seconds)	I/O Wait (seconds)	CEAD (seconds)
5230	119	4302

Conclusions

Results were presented for three different models provided by GM covering solution sequences 103, 108, and 107. As these examples are representative of the current NVH work load at GM, the ability to run these very large problems both in serial and parallel mode, illustrating both capacity and capability, demonstrate the suitability of the SV1 for NVH analysis in a production environment. The DMAP level frequency segmented parallelization approach requires a robust queuing system to provide efficient use of hardware resources.

Major improvements in performance were demonstrated for a collection of real life large scale problems provided by General Motors. Results clearly indicate that the Cray SV1 is a robust cost effective NVH engine well suited for a production environment in the automotive industry.

References

- [1] The 42nd L. Ray Buckindale Lecture. CAE methods and their application to truck design. Technical Report SP-1310, Ford Motor Company, 1997.
- [2] W. Nack. Brake squeal analysis by finite elements and comparisons to dyno results. In Proc. ASME Design Engineering Technical Conference, Las Vegas, Sept. 12-15, 1999.
- [3] N. Wichmann, L. Stern, K. Maschhoff. Performance improvements for NVH analysis on Cray SV1 Computers. ISATA 2000, Dublin, Ireland, Sept. 25-27, 2000.
- [4] L. Komzsik. Numerical Methods User's Guide, V70.5. The MacNeal-Schwendler Corporation, 1998.