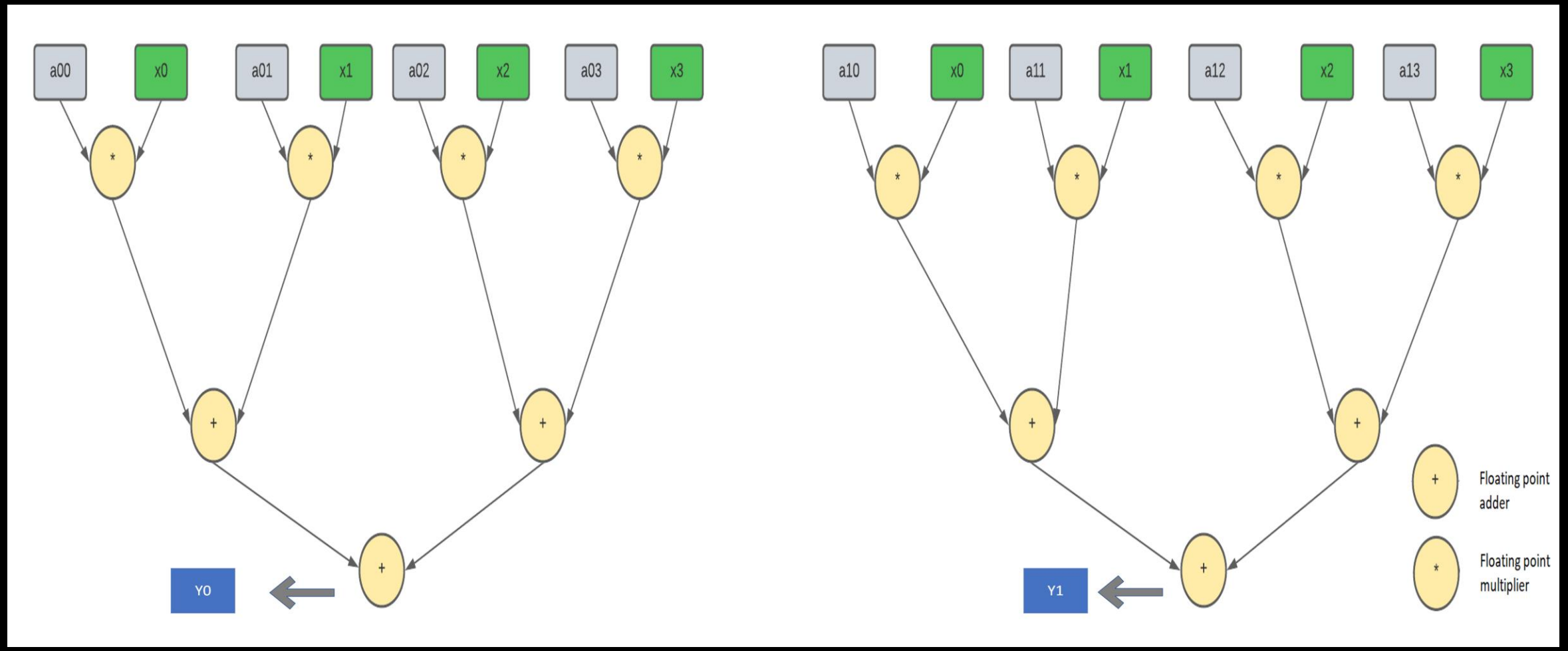
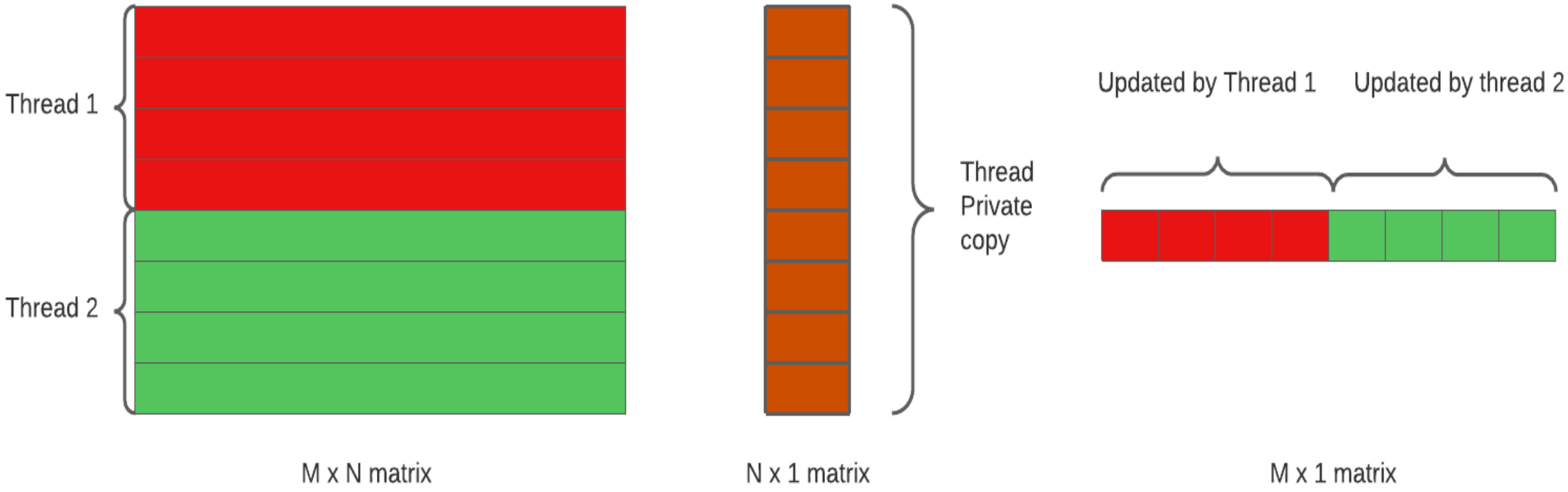
Multithreading SGEMV

Harihara Sudhan S, Software Design Engineer 2
Bhaskar Nallani, Senior Member Technical Staff

Var 1 SGEMV

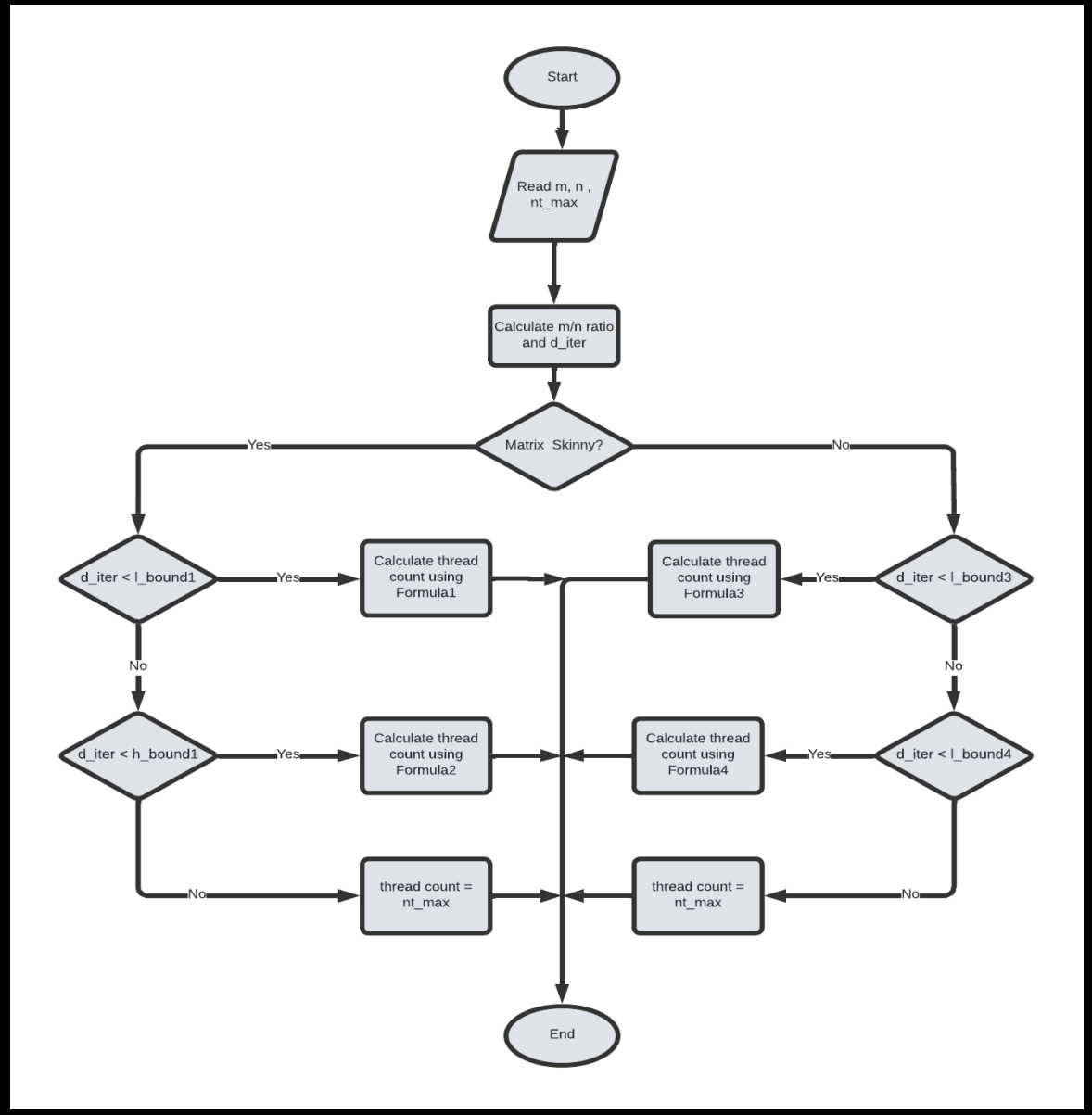


Var 1 SGEMV Multithreading

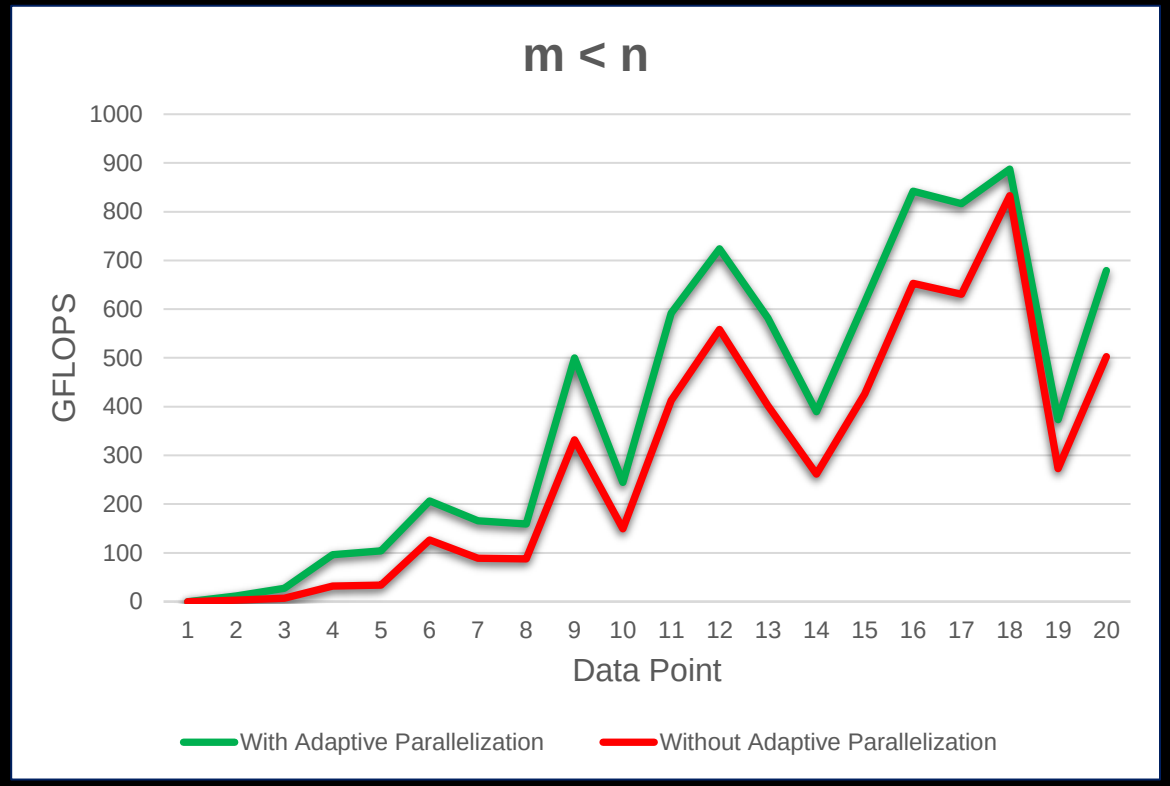


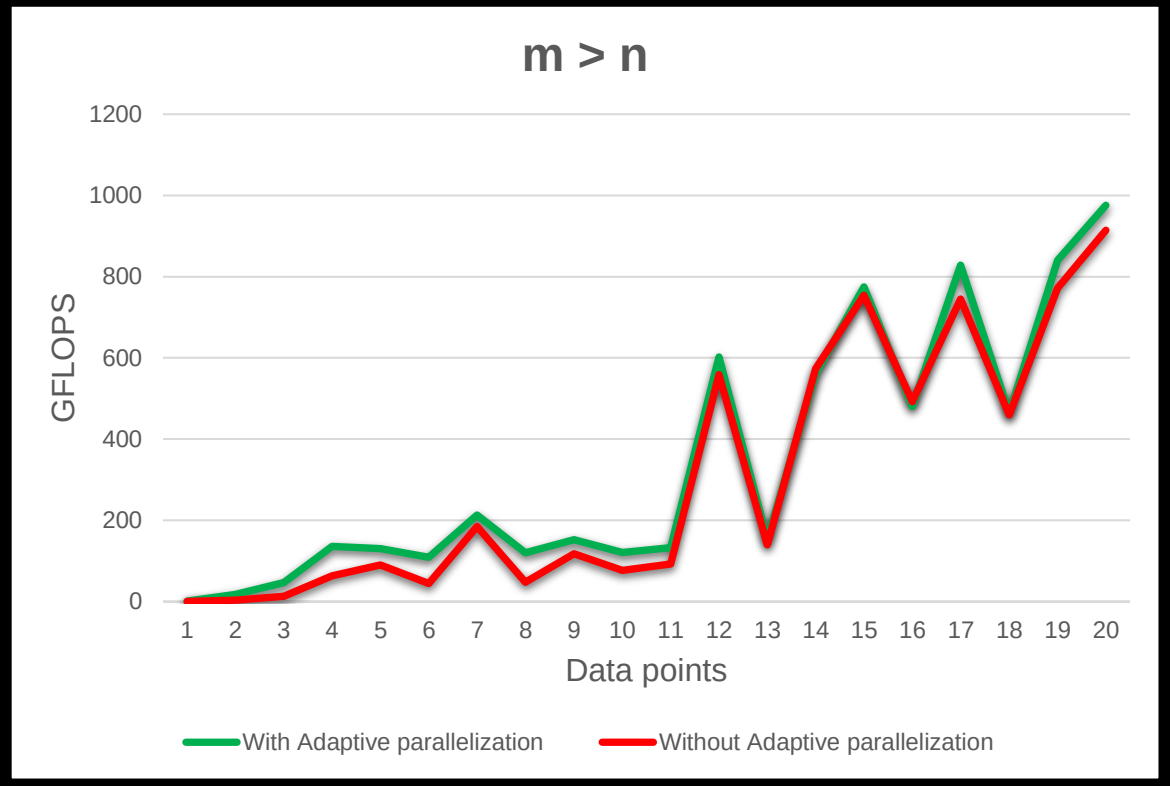
Adaptive Parallelization

- For matrix size was large (matrix size greater than 80% of L2 cache capacity) the peak performance achieved when the maximum number of threads spawned
  - Reduced the input window for which the threading logic was to be derived
- Classify based on the shape of the matrix.
  - Tall-and-skinny
  - Other (includes fat matrices and square)
- Classify based on size
- Use empirical equations derived from these regression models to calculate optimal number of threads



Performance





Input type	With Adaptive Parallelization	Without Adaptive Parallelization
m < n	163.94	96.26
m > n	172.88	108.63

Other GEMV Optimizations

- Standalone kernel for SGEMV var 1 multithreading
- DGEMV (both var 1 and var 2)
  - Packing of input matrix when stride is non unit
  - Aim – Performance improvement

Thank you

Questions and suggestions?

Copyright and disclaimer

- The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. The information contained herein is subject to change and may be rendered inaccurate releases, for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. Any computer system has risks of security vulnerabilities that cannot be completely prevented or mitigated. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

- THIS INFORMATION IS PROVIDED 'AS IS." AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS, OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION. AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY, OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY RELIANCE, DIRECT, INDIRECT, SPECIAL, OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.


AMD