FT-LA

As supercomputers grow ever larger in scale, the Mean Time to Failure becomes shorter and shorter, making the complete and successful execution of complex applications more and more difficult. **FT-LA delivers a new approach, utilizing Algorithm-Based Fault Tolerance (ABFT), to help factorization algorithms survive fail-stop failures.** The FT-LA software package extends ScaLAPACK with ABFT routines, and in sharp contrast with legacy checkpoint-based approaches, ABFT does not incur I/O overhead, and promises a much more scalable protection scheme.

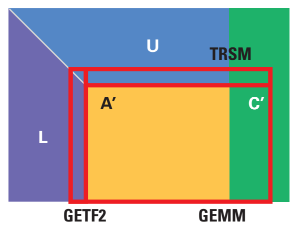
ABFT THE IDEA

PROTECTION

Matrix protected by block row ***checksum***

The ***algorithm*** updates both the ***trailing matrix*** AND the ***checksums***

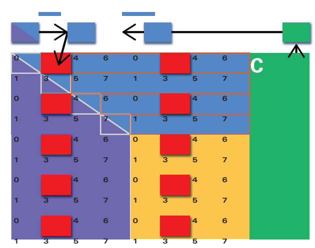
$$C_i = \sum_j A_{ij}$$



RECOVERY

Missing blocks reconstructed by inverting the checksum operation

$$A_{ij} = C_i - \sum_{k \neq j} A_{ik}$$



Process grid: $p \times q$

F: ***simultaneous*** failures tolerated

Usually $F << q$;
Overheads in F/q

Protection against 2 faults on 192x192 processes => 1% overhead

Computation

$$O\left(\frac{F}{q} \times M^3\right)$$

Flops for the checksum update

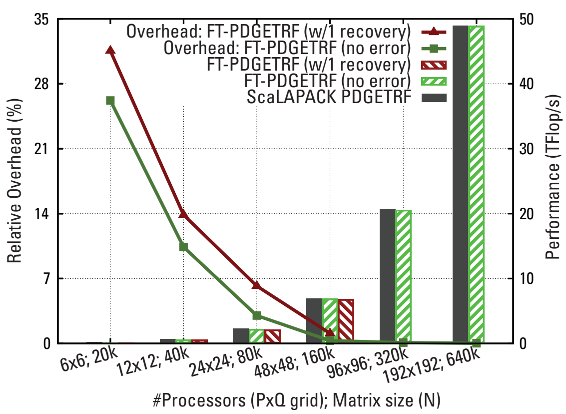
Memory

$$O\left(\frac{F}{q} \times M \times N\right)$$

Matrix is extended with $2F$ columns every q columns

Protection cost is inversely proportional to machine scale!

PERFORMANCE ON KRAKEN



- Cost of ABFT comes only from extra flops (to update checksums) and extra storage
- Cost decreases with machine scale (divided by Q when using $P \times Q$ processes)

FUNCTIONALITY	COVERAGE
Linear Systems of Equations	Cholesky, LU
Least Squares	QR (with protection of the upper and lower factors)
FEATURES	
Covering four precisions: double complex, single complex, double real, single real (ZCDS)	
Deploys on MPI FT draft (ULFM), or with “Checkpoint-on-failure”	
Allows toleration of permanent crashes	
WORK IN PROGRESS	
Hessenberg Reduction, Soft (silent) Errors	

ICL UT INNOVATIVE COMPUTING LABORATORY
THE UNIVERSITY of TENNESSEE