Quantifying Checkpoint Efficiency

John T. Daly, HPC-4

Advanced Simulation and Computing (ASC) applications implement fault tolerance using a simple checkpoint rollback strategy that periodically saves a known good state of a calculation by writing a dump file. One would like to be able to quantify the effectiveness of this form of fault tolerance, particularly in light of decreasing application mean time to interrupt (MTTI) associated with high-performance computing (HPC) systems of increasing complexity and numbers of parts.

If *solve time* is defined as time spent performing productive computation, and input/output (I/O), and run time as solve time plus time spent in restart, rework, and writing dump files, then it follows that the *checkpoint efficiency* can be accurately quantified in terms of three parameters: *restart overhead (R), dump time (δ), application MTTI (M), and checkpoint interval (t_c), as follows [1]:*

$$\text{Checkpoint Efficiency} \equiv \frac{T_s}{T_r} = e^{-\frac{R}{M}}\left(\frac{\lambda - \frac{1}{2}\Delta^2}{e^{\lambda} - 1}\right)$$

$$\text{where} \quad \lambda = \frac{t_c}{M}, \ \Delta = \sqrt{\frac{2\delta}{M}} \ .$$

By assuming that interrupts arrive randomly and are exponentially distributed [2], it is known that a solution for the optimum checkpoint interval exists in terms of dump time and application MTTI. It can be approximated to highest order when Δ is small [3], or it can be expressed analytically in terms of the Lambert W-function and approximated quite accurately in terms of a perturbation series [4]. The checkpoint efficiency as quantified above can be used to demonstrate that checkpoint schemes are not particularly sensitive to one's estimate of the application MTTI, as illustrated by Fig. 1. This property is important because in practice the application MTTI may be quite difficult to estimate accurately.

Since computing checkpoint efficiency in terms of an exact solution for the optimum checkpoint interval can be tedious, a simple approximation of λ that is accurate for both large and small Δ would be desirable. Such an approximation may be constructed using the notion of asymptotic matching as follows [1]:

$$\lambda \approx \Delta + \tfrac{1}{2}\Delta^2 \ \Rightarrow \text{Checkpoint Efficiency}$$

$$\approx e^{-\frac{R}{M}}\left(\frac{\Delta}{e^{\Delta + \frac{1}{2}\Delta^2} - 1}\right) \ .$$

Figure 2 demonstrates the accuracy of this approximation. It also clearly illustrates that once a 90–95% efficiency has been attained, further improvements in the ratio of reliability to dump time are of limited value to a checkpointing application. This is an important design consideration for future HPC systems.

[1] J.T. Daly, NECDC Proceedings (2007).
[2] J. T. Daly, LA-UR-06-8089 (2006).
[3] J. W. Young, *Comm. of the ACM* **17**, 530 (1974).
[4] J. T. Daly, "A Higher Order Estimate of the Optimum Checkpoint Interval for Restart Dumps," Future Generation Computer Systems **22**, 300 (2006).

For more information contact John Daly at jtd@lanl.gov.

Funding Acknowledgements
This research was supported by the NNSA tri-Lab Advanced Simulation and Computing Program.

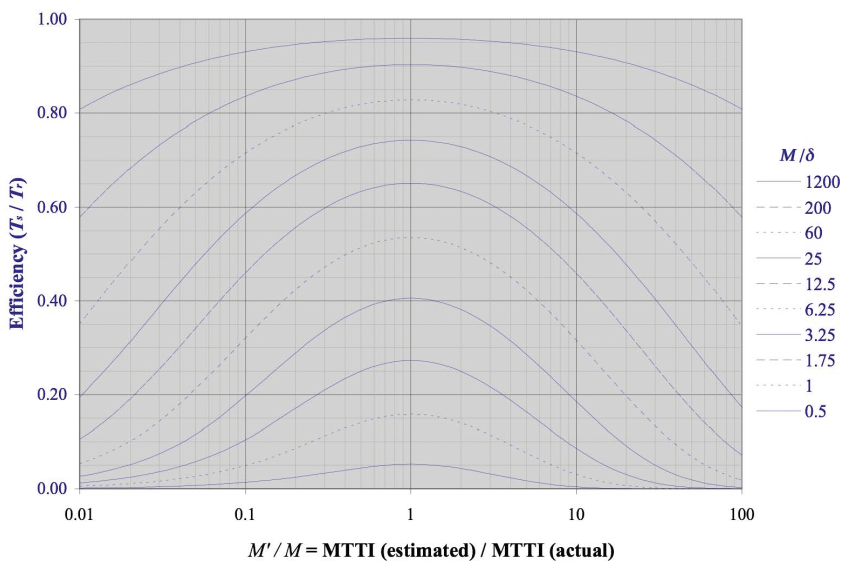

Fig. 1.
Checkpoint efficiency as a function of the accuracy of the estimated application MTTI based on the exact solution of the optimum checkpoint interval and assuming that restart overhead is negligible.

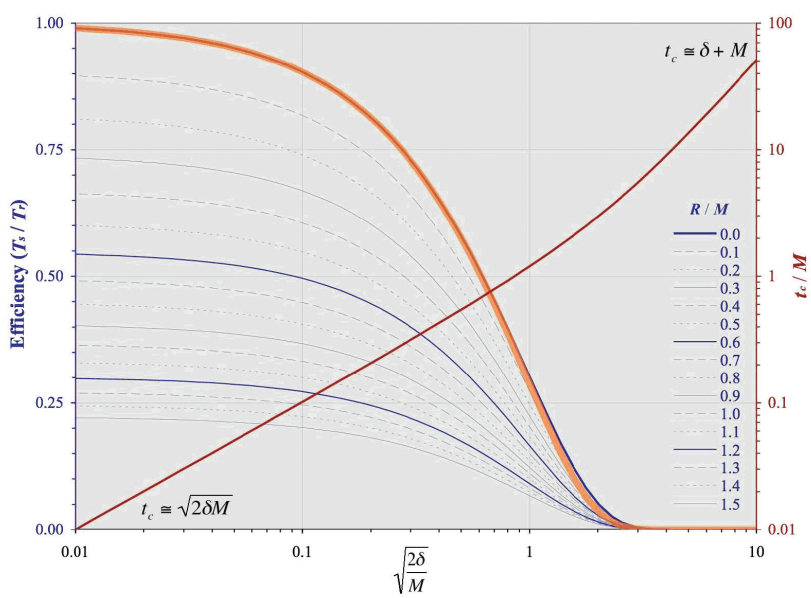

Fig. 2.
Checkpoint efficiency is plotted in blue. It is based on an exact solution of the optimum checkpoint interval shown in red. The efficiency based on asymptotic matching is plotted in orange for R/M =0.