**Advancing Computational Electromagnetics Research though Benchmarking**

Ali E. Yılmaz[(1)]
(1) The University of Texas at Austin, Austin, TX, 78712

Benchmarking is a familiar idea to many in the context of computer business and marketing, where system performances have long been compared using standardized benchmarks. The focus of such benchmarking is more on purchasing decisions and competitive aspects rather than research, development, or educational aspects of the process. A different use of benchmarking is frequently mentioned in the context of software engineering, where benchmarks that are created and used by a technical research community are often celebrated and considered to encourage scientific progress. The theory of benchmarking developed in (S. E. Sim, S. Easterbrook, and R. C. Holt, Proc. IEEE Comp. Soc. 25th Int. Conf. Software Eng., pp. 74-83, 2003) identifies three prerequisites that must exist within a research discipline before benchmarking can be used to make advances: a minimum level of maturity, a tradition of comparing research results, and an ethos of collaboration. This article asserts that computational electromagnetics (CEM) research currently meets these conditions and benchmarking is a fertile topic of inquiry in CEM.

Today benchmarking in CEM research is commonly understood as the task of using a (new) algorithm to solve sample scattering problems (by those creating and implementing the algorithm) and comparing the newly computed values of certain quantities of interest (e.g., radar cross section) to theoretical, experimental, or previously computed references. The primary goal of this mode of benchmarking is to obtain empirical evidence that validates (and can potentially falsify) theoretically expected features of an algorithm, e.g., its error convergence rate or asymptotic cost scaling with respect to degrees of freedom (J. W. Massey, A. Menshov, and A. E. Yilmaz, URSI NRSM, Jan. 2017). Beyond this role, however, benchmarking can also be used for other purposes:

1. Benchmarking is a systematic tool that can be used to combat the ubiquity of error. It is closely related to the increasingly urgent calls for reproducible scientific and engineering software development but does not place undue burdens of (perfect) replication on researchers.
2. Benchmarking can be used to combat the hazards of specialization in research: It can inform others about important problems, about the current state of computational systems for solving these problems, help researchers keep up with advances, help scientists keep an open mind, and lower barriers to entry of new ideas, researchers, and tools.
3. Benchmarking can reduce importance of subjective factors when judging simulation tools, increase credibility of claims made by computational scientists and engineers, and fortify their intellectual and scientific integrity.
4. Benchmarking can highlight open problems that demand new solution methods, identify weaknesses in existing computational systems, and inspire research to address these weaknesses.

While existing benchmark suites rarely contain all the ingredients necessary to perform such benchmarking, the recently developed Austin BioEM benchmark suite (J. W. Massey et al., URSI Int. Symp. Electromagnetic Theory, Aug. 2016) indicates that there are many emerging possibilities. At the conference, some of the advantages and pitfalls of public benchmarking will discussed.