

Introduction

- As HPC applications become more complex, development through accretion is untenable and the methods and tools of software productivity are necessary.

- We assess the tools and methods of 15 applications from ECP, INCITE, and NSF programs that span a variety of scientific domains.

- Information is gathered from:
 - ECP team interviews,
 - web materials: documentation and presentations,
 - analysis of code repositories (see also Kanika Sood's poster)

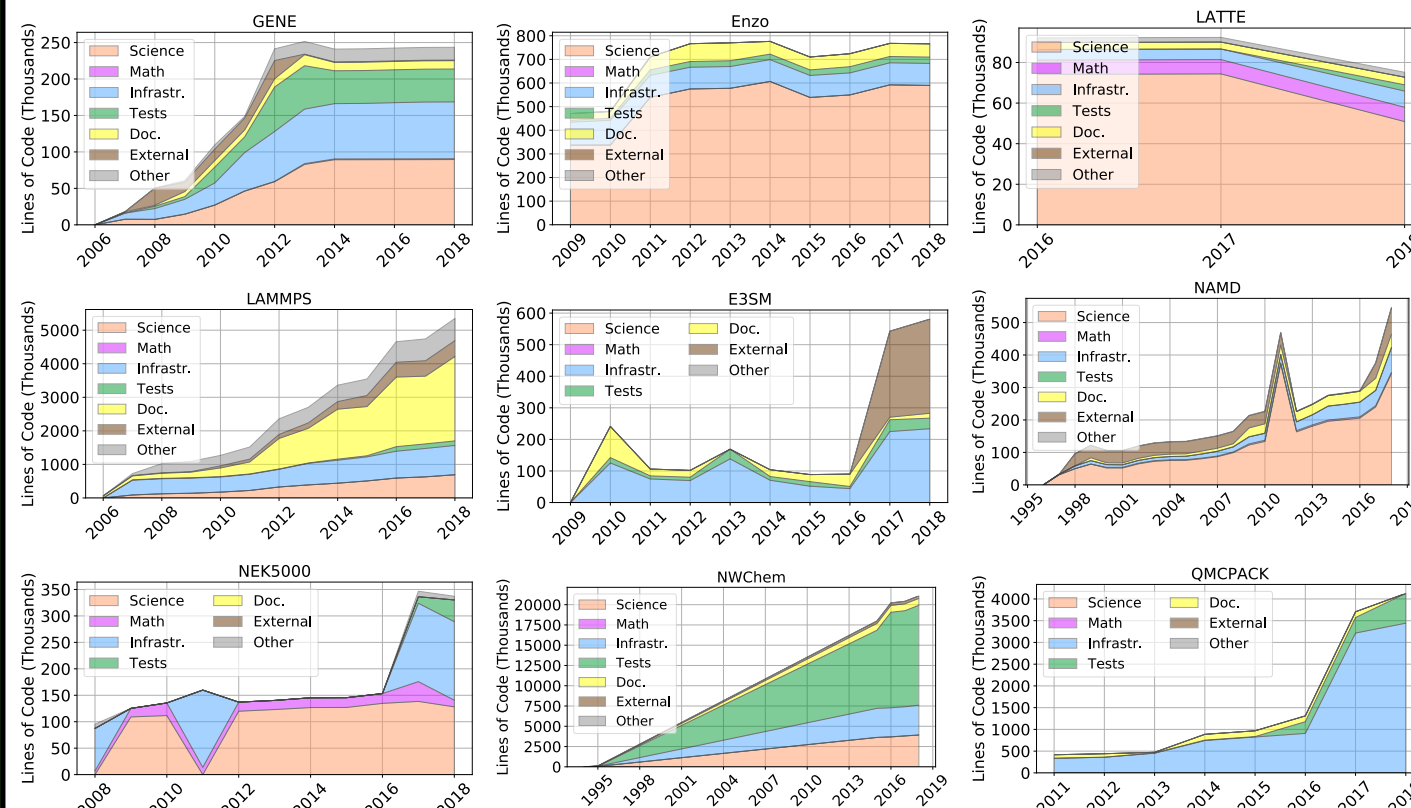
1. Code overview

- The table summarizes code longevity, team size, version control and testing practices, and the user communities and domains served.

Code	Year opened	Repo	Lifetime	Capabilites	Version	Capability growth	Size growth	Team Size	Interdisc.	Testing	CI	Domains	Community
Amber	NA		1975	molecular dynamics	Amber18	methods		40 principal contributors		likely automated	likely yes	several	1000+
Cactus/ET	1997	git	1997	general relativity		AMR, scaling	NA	9 core (94 members)	yes	automated	no	2	large
E3SM	2018	git	2012	Earth systems	1.0	NA	NA	20	yes	automated	manually	5	400+
Enzo	2004	bitbucket	1996	multiphysics	2.5 released	methods	1.249	12 core + 25 contrib.	no	likely automated	no	>6	1000+
FLASH	2000	svn & git	1998	multiphysics	4.5 released	physics & methods	> 5 times	~6 core + contrib.	yes	automated	in ver. 5	1	100, larger for Cactus
GENE	open	git	1999	plasma physics	11 released	performance	4.7	14 core + contrib.	no	manual	no	3	7000+
HACC	NA		2014	cosmology	closed			11 core	no	likely manual	no	1	70k+ downloads
LAMMPS	2004	git	1995	molecular dynamics	22Aug2018	methods	1.97	~4 core	no	automated	yes	1	1k downloads
LATTE	2016	git	2007	molecular dynamics			1.035	5-6	yes		yes	>6	200
MARBL	NA	git	2014	multiphysics	closed			~20	no	automated	yes	1	Core group only
NEK5000	2009	git	1999	reactive CFD	limited	performance	2.658	~10	no	manual	no	1	236
NAMD	1998	git	1995	molecular dynamics	2.13 released	methods		4 core	yes	likely automated	no	1	~10
NWChem	2010	svn & git	1996	chemistry	6.8.1 released	performance	1.010	5 core + 17 active	yes	likely automated	some	3	DOE climate community
QMCPACK	2003	git	1998	chem & quantum phy.	3.5 released	methods	1.128	~5 core	no	automated	yes	>3	NA
XGC	open	bitbucket	2002	plasma physics	3 released			4 core	no	manual	not currently		

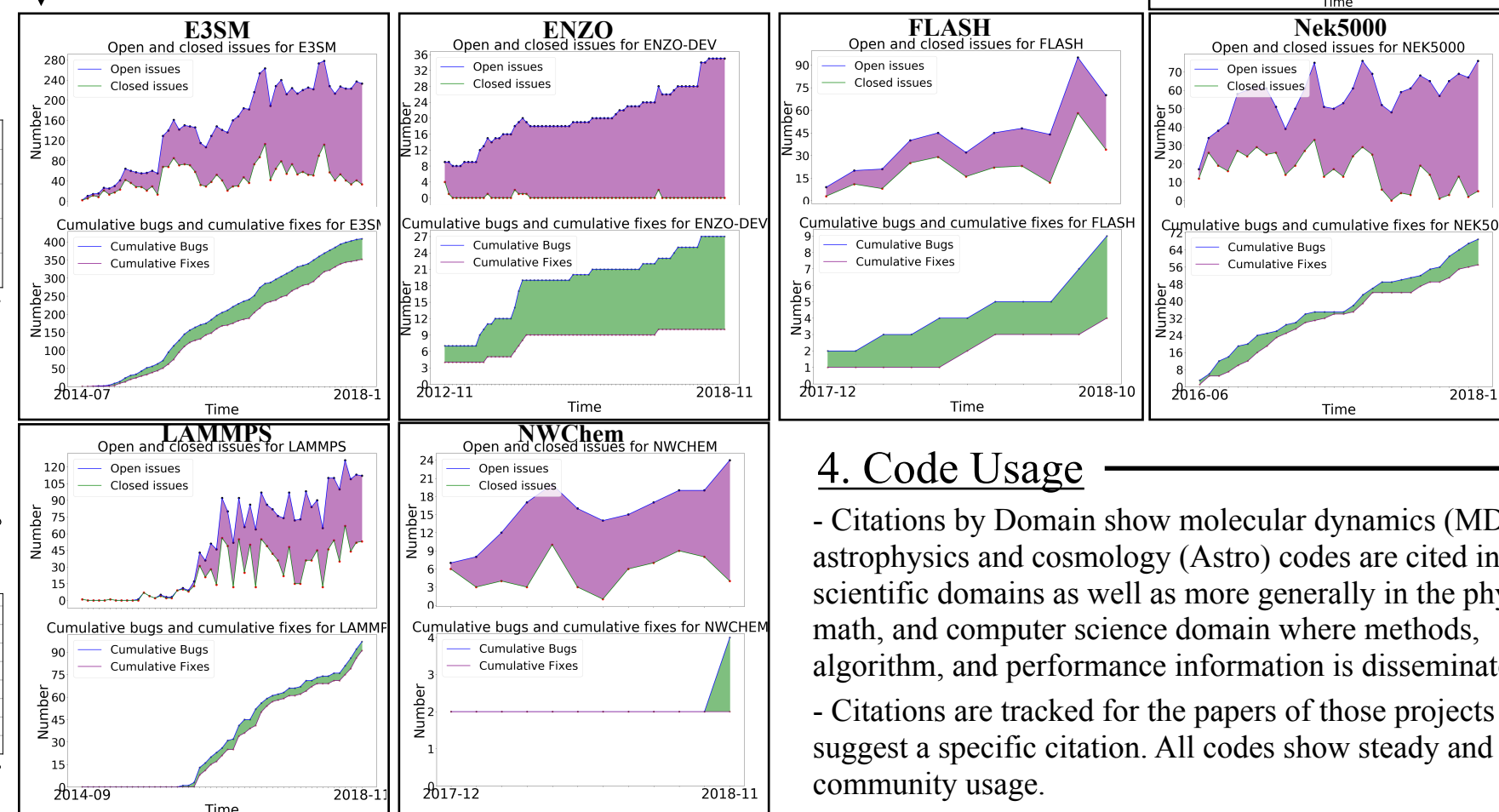
2. Code Size

- Lines of code (LOC) divided into Science, Math, Infrastructure, Testing, Documentation, External, and Other and plotted over time. Some LOC counts are steady while the majority are increasing with testing and documentation providing a significant portion of the increase.



3. Code Issue/Bug Tracking

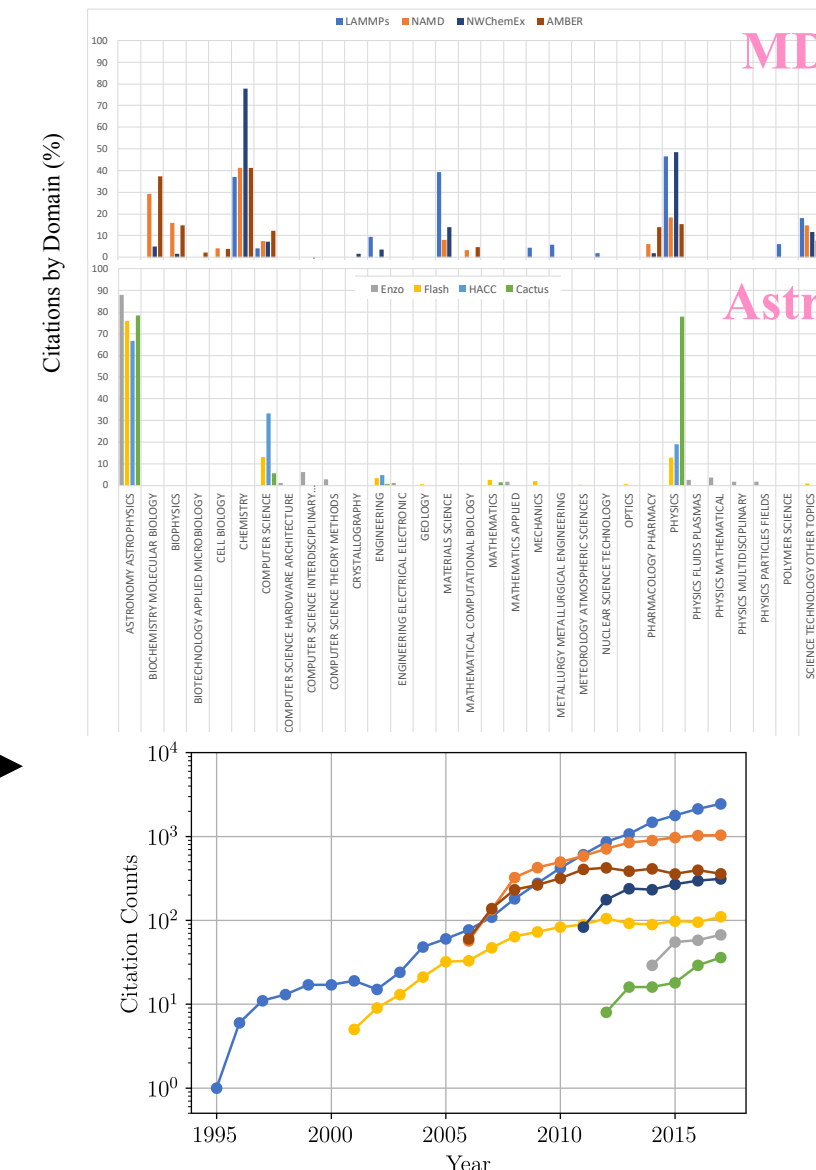
From repository analysis, "river" plots show the opening and closing of issues and bugs. The bugs "river" generally remains narrow indicating prioritization of effort to resolve bugs while the issues "river" tends to widen.



4. Code Usage

- Citations by Domain show molecular dynamics (MD) and astrophysics and cosmology (Astro) codes are cited in their scientific domains as well as more generally in the physics, math, and computer science domain where methods, algorithm, and performance information is disseminated.

- Citations are tracked for the papers of those projects that suggest a specific citation. All codes show steady and stable community usage.



Takeaways

1. Code survey shows that many projects are moving toward open source distribution with implementations of version control and testing practices.
2. Tracking code size helps teams consider project composition and aid in planning of major efforts like refactoring.
3. Analysis of bug and issue tracking help teams plan how resources are allocated to address the demands of users.
4. Trends in citation statistics provide important metrics for usability and domain breadth that is useful for long term development and funding plans.