



# SymPy: symbolic computing in Python

Aaron Meurer<sup>1</sup>, Christopher P. Smith<sup>2</sup>, Mateusz Paprocki<sup>3</sup>, Ondřej Čertík<sup>4</sup>, Sergey B. Kirpichev<sup>5</sup>, Matthew Rocklin<sup>3</sup>, AMiT Kumar<sup>6</sup>, Sergiu Ivanov<sup>7</sup>, Jason K. Moore<sup>8</sup>, Sartaj Singh<sup>9</sup>, Thilina Rathnayake<sup>10</sup>, Sean Vig<sup>11</sup>, Brian E. Granger<sup>12</sup>, Richard P. Muller<sup>13</sup>, Francesco Bonazzi<sup>14</sup>, Harsh Gupta<sup>15</sup>, Shivam Vats<sup>15</sup>, Fredrik Johansson<sup>16</sup>, Fabian Pedregosa<sup>17</sup>, Matthew J. Curry<sup>18,19,20</sup>, Andy R. Terrel<sup>21,22</sup>, Štěpán Roučka<sup>23</sup>, Ashutosh Saboo<sup>24</sup>, Isuru Fernando<sup>10</sup>, Sumith Kulal<sup>25</sup>, Robert Cimrman<sup>26</sup> and Anthony Scopatz<sup>1</sup>

- <sup>1</sup> Department of Mechanical Engineering, University of South Carolina, Columbia, SC, United States  
<sup>2</sup> Polar Semiconductor, Inc., Bloomington, MN, United States  
<sup>3</sup> Continuum Analytics, Inc., Austin, TX, United States  
<sup>4</sup> Los Alamos National Laboratory, Los Alamos, NM, United States  
<sup>5</sup> Faculty of Physics, Moscow State University, Moscow, Russia  
<sup>6</sup> Department of Applied Mathematics, Delhi Technological University, New Delhi, India  
<sup>7</sup> Université Paris Est Créteil, Créteil, France  
<sup>8</sup> Mechanical and Aerospace Engineering, University of California, Davis, CA, United States  
<sup>9</sup> Mathematical Sciences, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India  
<sup>10</sup> Department of Computer Science and Engineering, University of Moratuwa, Katubedda, Moratuwa, Sri Lanka  
<sup>11</sup> University of Illinois at Urbana-Champaign, Urbana, IL, United States  
<sup>12</sup> California Polytechnic State University, San Luis Obispo, CA, United States  
<sup>13</sup> Center for Computing Research, Sandia National Laboratories, Albuquerque, NM, United States  
<sup>14</sup> Department of Theory and Bio-Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany  
<sup>15</sup> Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India  
<sup>16</sup> INRIA Bordeaux-Sud-Ouest—LFANT project-team, Talence, France  
<sup>17</sup> INRIA—SIERRA project-team, Paris, France  
<sup>18</sup> Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, United States  
<sup>19</sup> Center for Quantum Information and Control, University of New Mexico, Albuquerque, NM, United States  
<sup>20</sup> Sandia National Laboratories, Albuquerque, NM, United States  
<sup>21</sup> Fashion Metric, Inc, Austin, TX, United States  
<sup>22</sup> NumFOCUS, Austin, TX, United States  
<sup>23</sup> Department of Surface and Plasma Science, Faculty of Mathematics and Physics, Charles University in Prague, Praha, Czech Republic  
<sup>24</sup> Department of Computer Science, Department of Mathematics, Birla Institute of Technology and Science, Goa, India  
<sup>25</sup> Indian Institute of Technology Bombay, Mumbai, Maharashtra, India  
<sup>26</sup> New Technologies—Research Centre, University of West Bohemia, Plzeň, Czech Republic

Submitted 22 June 2016  
Accepted 21 November 2016  
Published 02 January 2017

Corresponding author  
Aaron Meurer, [asmeurer@gmail.com](mailto:asmeurer@gmail.com)

Academic editor  
Nick Higham

Additional Information and  
Declarations can be found on  
page 22

DOI 10.7717/peerj-cs.103

© Copyright  
2017 Meurer et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

## ABSTRACT

SymPy is an open source computer algebra system written in pure Python. It is built with a focus on extensibility and ease of use, through both interactive and programmatic applications. These characteristics have led SymPy to become a popular symbolic library for the scientific Python ecosystem. This paper presents the architecture of SymPy, a description of its features, and a discussion of select submodules. The supplementary material provide additional examples and further outline details of the architecture and features of SymPy.

**Subjects** Scientific Computing and Simulation, Software Engineering

**Keywords** Python, Computer algebra system, Symbolics

## INTRODUCTION

SymPy is a full featured computer algebra system (CAS) written in the Python (Lutz, 2013) programming language. It is free and open source software, licensed under the 3-clause BSD license (Rosen, 2005). The SymPy project was started by Ondřej Čertík in 2005, and it has since grown to over 500 contributors. Currently, SymPy is developed on GitHub using a bazaar community model (Raymond, 1999). The accessibility of the codebase and the open community model allow SymPy to rapidly respond to the needs of users and developers.

Python is a dynamically typed programming language that has a focus on ease of use and readability.<sup>1</sup> Due in part to this focus, it has become a popular language for scientific computing and data science, with a broad ecosystem of libraries (Oliphant, 2007). SymPy is itself used as a dependency by many libraries and tools to support research within a variety of domains, such as SageMath (The Sage Developers, 2016) (pure and applied mathematics), yt (Turk et al., 2011) (astronomy and astrophysics), PyDy (Gede et al., 2013) (multibody dynamics), and SfePy (Cimrman, 2014) (finite elements).

Unlike many CAS's, SymPy does not invent its own programming language. Python itself is used both for the internal implementation and end user interaction. By using the operator overloading functionality of Python, SymPy follows the embedded domain specific language paradigm proposed by Hudak (1998). The exclusive usage of a single programming language makes it easier for people already familiar with that language to use or develop SymPy. Simultaneously, it enables developers to focus on mathematics, rather than language design. SymPy version 1.0 officially supports Python 2.6, 2.7 and 3.2–3.5.

SymPy is designed with a strong focus on usability as a library. Extensibility is important in its application program interface (API) design. Thus, SymPy makes no attempt to extend the Python language itself. The goal is for users of SymPy to be able to include SymPy alongside other Python libraries in their workflow, whether that be in an interactive environment or as a programmatic part in a larger system.

Being a library, SymPy does not have a built-in graphical user interface (GUI). However, SymPy exposes a rich interactive display system, and supports registering display formatters with Jupyter (Kluyver et al., 2016) frontends, including the Notebook and Qt Console, which will render SymPy expressions using MathJax (Cervone, 2012) or  $\text{\LaTeX}$ .

The remainder of this paper discusses key components of the SymPy library. Section 'Overview of capabilities' enumerates the features of SymPy and takes a closer look at some of the important ones. The Section 'Numerics' looks at the numerical features of SymPy and its dependency library, mpmath. Section 'Physics Submodule' looks at the domain specific physics submodules for performing symbolic and numerical calculations in classical mechanics and quantum mechanics. Section 'Architecture' discusses the architecture of SymPy. Section 'Projects that Depend on SymPy' looks at a selection of packages that depend on SymPy. Conclusions and future directions for SymPy are given in 'Conclusion and Future Work'. All examples in this paper use SymPy version 1.0 and mpmath version 0.19.

Additionally, the [Supplemental Information 1](#) takes a deeper look at a few SymPy topics. [Section S1](#) discusses the Gruntz algorithm, which SymPy uses to calculate symbolic limits.

<sup>1</sup>This paper assumes a moderate familiarity with the Python programming language.

Sections S2–S9 of the supplement discuss the series, logic, Diophantine equations, sets, statistics, category theory, tensor, and numerical simplification submodules of SymPy, respectively. Section S10 provides additional examples for topics discussed in the main paper. Section S11 discusses the SymPy Gamma project. Finally, Section S12 of the supplement contains a brief comparison of SymPy with Wolfram Mathematica.

<sup>2</sup> `import *` has been used here to aid the readability of the paper, but is best to avoid such wildcard import statements in production code, as they make it unclear which names are present in the namespace. Furthermore, imported names could clash with already existing imports from another package. For example, SymPy, the standard Python math library, and NumPy all define the `exp` function, but only the SymPy one will work with SymPy symbolic expressions.

<sup>3</sup> The three greater-than signs denote the user input for the Python interactive session, with the result, if there is one, shown on the next line.

The following statement imports all SymPy functions into the global Python namespace.<sup>2</sup> From here on, all examples in this paper assume that this statement has been executed<sup>3</sup>:

```
>>> from sympy import *
```

All the examples in this paper can be tested on [SymPy Live](#), an online Python shell that uses the Google App Engine ([Ciurana, 2009](#)) to execute SymPy code. SymPy Live is also integrated into the SymPy documentation at <http://docs.sympy.org>.

## OVERVIEW OF CAPABILITIES

This section gives a basic introduction of SymPy, and lists its features. A few features—assumptions, simplification, calculus, polynomials, printers, solvers, and matrices—are core components of SymPy and are discussed in depth. Many other features are discussed in depth in the [Supplemental Information 1](#).

### Basic usage

Symbolic variables, called symbols, must be defined and assigned to Python variables before they can be used. This is typically done through the `symbols` function, which may create multiple symbols in a single function call. For instance,

```
>>> x, y, z = symbols('x y z')
```

creates three symbols representing variables named  $x$ ,  $y$ , and  $z$ . In this particular instance, these symbols are all assigned to Python variables of the same name. However, the user is free to assign them to different Python variables, while representing the same symbol, such as `a, b, c = symbols('x y z')`. In order to minimize potential confusion, though, all examples in this paper will assume that the symbols  $x$ ,  $y$ , and  $z$  have been assigned to Python variables identical to their symbolic names.

Expressions are created from symbols using Python's mathematical syntax. For instance, the following Python code creates the expression  $(x^2 - 2x + 3)/y$ . Note that the expression remains unevaluated: it is represented symbolically.

```
>>> (x**2 - 2*x + 3)/y
(x**2 - 2*x + 3)/y
```

### List of features

Although SymPy's extensive feature set cannot be covered in depth in this paper, bedrock areas, that is, those areas that are used throughout the library, are discussed in their own subsections below. Additionally, [Table 1](#) gives a compact listing of all major capabilities present in the SymPy codebase. This grants a sampling from the breadth of topics and application domains that SymPy services. Unless stated otherwise, all features noted in [Table 1](#) are symbolic in nature. Numeric features are discussed in Section 'Numerics.'

**Table 1** SymPy features and descriptions.

Feature (submodules)	Description
Calculus (sympy.core, sympy.calculus, sympy.integrals, sympy.series)	Algorithms for computing derivatives, integrals, and limits.
Category Theory (sympy.categories)	Representation of objects, morphisms, and diagrams. Tools for drawing diagrams with Xy-pic ( <a href="#">Rose, 1999</a> ).
Code Generation (sympy.printing, sympy.codegen)	Generation of compilable and executable code in a variety of different programming languages from expressions directly. Target languages include C, Fortran, Julia, JavaScript, Mathematica, MATLAB and Octave, Python, and Theano.
Combinatorics & Group Theory (sympy.combinatorics)	Permutations, combinations, partitions, subsets, various permutation groups (such as polyhedral, Rubik, symmetric, and others), Gray codes ( <a href="#">Nijenhuis &amp; Wilf, 1978</a> ), and Prufer sequences ( <a href="#">Biggs, Lloyd &amp; Wilson, 1976</a> ).
Concrete Math (sympy.concrete)	Summation, products, tools for determining whether summation and product expressions are convergent, absolutely convergent, hypergeometric, and for determining other properties; computation of Gosper's normal form ( <a href="#">Petkovšek, Wilf &amp; Zeilberger, 1997</a> ) for two univariate polynomials.
Cryptography (sympy.crypto)	Block and stream ciphers, including shift, Affine, substitution, Vigenère's, Hill's, bifid, RSA, Kid RSA, linear-feedback shift registers, and Elgamal encryption.
Differential Geometry (sympy.diffgeom)	Representations of manifolds, metrics, tensor products, and coordinate systems in Riemannian and pseudo-Riemannian geometries ( <a href="#">Sussman &amp; Wisdom, 2013</a> ).
Geometry (sympy.geometry)	Representations of 2D geometrical entities, such as lines and circles. Enables queries on these entities, such as asking the area of an ellipse, checking for collinearity of a set of points, or finding the intersection between objects.
Lie Algebras (sympy.liealgebras)	Representations of Lie algebras and root systems.
Logic (sympy.logic)	Boolean expressions, equivalence testing, satisfiability, and normal forms.
Matrices (sympy.matrices)	Tools for creating matrices of symbols and expressions. Both sparse and dense representations, as well as symbolic linear algebraic operations (e.g., inversion and factorization), are supported.
Matrix Expressions (sympy.matrices.expressions)	Matrices with symbolic dimensions (unspecified entries). Block matrices.
Number Theory (sympy.ntheory)	Prime number generation, primality testing, integer factorization, continued fractions, Egyptian fractions, modular arithmetic, quadratic residues, partitions, binomial and multinomial coefficients, prime number tools, hexadecimal digits of $\pi$ , and integer factorization.
Plotting (sympy.plotting)	Hooks for visualizing expressions via matplotlib ( <a href="#">Hunter, 2007</a> ) or as text drawings when lacking a graphical backend. 2D function plotting, 3D function plotting, and 2D implicit function plotting are supported.

(continued on next page)

Table 1 (continued)

Feature (submodules)	Description
Polynomials ( <code>sympy.polys</code> )	Polynomial algebras over various coefficient domains. Functionality ranges from simple operations (e.g., polynomial division) to advanced computations (e.g., Gröbner bases ( <i>Adams &amp; Loustau, 1994</i> ) and multivariate factorization over algebraic number domains).
Printing ( <code>sympy.printing</code> )	Functions for printing SymPy expressions in the terminal with ASCII or Unicode characters and converting SymPy expressions to $\text{\LaTeX}$ and MathML.
Quantum Mechanics ( <code>sympy.physics.quantum</code> )	Quantum states, bra–ket notation, operators, basis sets, representations, tensor products, inner products, outer products, commutators, anticommutators, and specific quantum system implementations.
Series ( <code>sympy.series</code> )	Series expansion, sequences, and limits of sequences. This includes Taylor, Laurent, and Puiseux series as well as special series, such as Fourier and formal power series.
Sets ( <code>sympy.sets</code> )	Representations of empty, finite, and infinite sets (including special sets such as the natural, integer, and complex numbers). Operations on sets such as union, intersection, Cartesian product, and building sets from other sets are supported.
Simplification ( <code>sympy.simplify</code> )	Functions for manipulating and simplifying expressions. Includes algorithms for simplifying hypergeometric functions, trigonometric expressions, rational functions, combinatorial functions, square root denesting, and common subexpression elimination.
Solvers ( <code>sympy.solvers</code> )	Functions for symbolically solving equations, systems of equations, both linear and non-linear, inequalities, ordinary differential equations, partial differential equations, Diophantine equations, and recurrence relations.
Special Functions ( <code>sympy.functions</code> )	Implementations of a number of well known special functions, including Dirac delta, Gamma, Beta, Gauss error functions, Fresnel integrals, Exponential integrals, Logarithmic integrals, Trigonometric integrals, Bessel, Hankel, Airy, B-spline, Riemann Zeta, Dirichlet eta, polylogarithm, Lerch transcendent, hypergeometric, elliptic integrals, Mathieu, Jacobi polynomials, Gegenbauer polynomial, Chebyshev polynomial, Legendre polynomial, Hermite polynomial, Laguerre polynomial, and spherical harmonic functions.
Statistics ( <code>sympy.stats</code> )	Support for a random variable type as well as the ability to declare this variable from prebuilt distribution functions such as Normal, Exponential, Coin, Die, and other custom distributions ( <i>Rocklin &amp; Terrel, 2012</i> ).
Tensors ( <code>sympy.tensor</code> )	Symbolic manipulation of indexed objects.
Vectors ( <code>sympy.vector</code> )	Basic operations on vectors and differential calculus with respect to 3D Cartesian coordinate systems.

<sup>4</sup>In SymPy,  $\sqrt{z}$  is defined on the usual principal branch with the branch cut along the negative real axis.

<sup>5</sup>SymPy assumes that two expressions  $A$  and  $B$  commute with each other multiplicatively, that is,  $A \cdot B = B \cdot A$ , unless they both have `commutative=False`. Many algorithms in SymPy require special consideration to work correctly with noncommutative products.

<sup>6</sup>For historical reasons, this algorithm is distinct from the `sympy.logic` submodule, which is discussed in [Section S3](#). SymPy also has an experimental assumptions system which stores facts separate from objects, and uses `sympy.logic` and a SAT solver for deduction. We will not discuss this system here.

## Assumptions

The assumptions system allows users to specify that symbols have certain common mathematical properties, such as being positive, imaginary, or integer. SymPy is careful to never perform simplifications on an expression unless the assumptions allow them. For instance, the simplification  $\sqrt{t^2} = t$  holds if  $t$  is nonnegative ( $t \geq 0$ ), but it does not hold for a general complex  $t$ .<sup>4</sup>

By default, SymPy performs all calculations assuming that symbols are complex valued. This assumption makes it easier to treat mathematical problems in full generality.

```
>>> t = Symbol('t')
>>> sqrt(t**2)
sqrt(t**2)
```

By assuming the most general case, that  $t$  is complex by default, SymPy avoids performing mathematically invalid operations. However, in many cases users will wish to simplify expressions containing terms like  $\sqrt{t^2}$ .

Assumptions are set on Symbol objects when they are created. For instance `Symbol('t', positive=True)` will create a symbol named  $t$  that is assumed to be positive.

```
>>> t = Symbol('t', positive=True)
>>> sqrt(t**2)
t
```

Some of the common assumptions are negative, real, nonpositive, integer, prime and commutative.<sup>5</sup> Assumptions on any SymPy object can be checked with the `is_` assumption attributes, like `t.is_positive`.

Assumptions are only needed to restrict a domain so that certain simplifications can be performed. They are not required to make the domain match the input of a function. For instance, one can create the object  $\sum_{n=0}^m f(n)$  as `Sum(f(n), (n, 0, m))` without setting `integer=True` when creating the Symbol object  $n$ .

The assumptions system additionally has deductive capabilities. The assumptions use a three-valued logic using the Python built in objects `True`, `False`, and `None`. Note that `False` is returned if the SymPy object doesn't or can't have the assumption. For example, both `I.is_real` and `I.is_prime` return `False` for the imaginary unit  $I$ .

`None` represents the “unknown” case. This could mean that given assumptions do not unambiguously specify the truth of an attribute. For instance, `Symbol('x', real=True).is_positive` will give `None` because a real symbol might be positive or negative. `None` could also mean that not enough is known or implemented to compute the given fact. For instance, `(pi + E).is_irrational` gives `None`—indeed, the rationality of  $\pi + e$  is an open problem in mathematics ([Lang, 1966](#)).

Basic implications between the facts are used to deduce assumptions. Deductions are made using the Rete algorithm ([Doorenbos, 1995](#)).<sup>6</sup> For instance, the assumptions system knows that being an integer implies being rational.

```
>>> i = Symbol('i', integer=True)
>>> i.is_rational
True
```



**Table 2** Some SymPy simplification functions.

expand	expand the expression
factor	factor a polynomial into irreducibles
collect	collect polynomial coefficients
cancel	rewrite a rational function as $p/q$ with common factors canceled
apart	compute the partial fraction decomposition of a rational function
trigsimp	simplify trigonometric expressions (Fu, Zhong & Zeng, 2006)
hyperexpand	expand hypergeometric functions (Roach, 1996; Roach, 1997)

Furthermore, expressions compute the assumptions on themselves based on the assumptions of their arguments. For instance, if  $x$  and  $y$  are both created with `positive=True`, then `(x + y).is_positive` will be `True` (whereas `(x - y).is_positive` will be `None`).

## Simplification

The generic way to simplify an expression is by calling the `simplify` function. It must be emphasized that simplification is not a rigorously defined mathematical operation (Moses, 1971). The `simplify` function applies several simplification routines along with heuristics to make the output expression “simple”.<sup>7</sup>

It is often preferable to apply more directed simplification functions. These apply very specific rules to the input expression and are typically able to make guarantees about the output. For instance, the `factor` function, given a polynomial with rational coefficients in several variables, is guaranteed to produce a factorization into irreducible factors. Table 2 lists common simplification functions.

Examples for these simplification functions can be found in Section S10.

## Calculus

SymPy provides all the basic operations of calculus, such as calculating limits, derivatives, integrals, or summations.

Limits are computed with the `limit` function, using the Gruntz algorithm (Gruntz, 1996) for computing symbolic limits and heuristics (a description of the Gruntz algorithm may be found in Section S1). For example, the following computes  $\lim_{x \rightarrow \infty} x \sin(\frac{1}{x}) = 1$ . Note that SymPy denotes  $\infty$  as `oo` (two lower case “o”s).

```
>>> limit(x*sin(1/x), x, oo)
1
```

As a more complex example, SymPy computes

$$\lim_{x \rightarrow 0} \left( 2e^{\frac{1-\cos(x)}{\sin(x)}} - 1 \right)^{\frac{\sinh(x)}{\tan^2(x)}} = e.$$

```
>>> limit((2*exp((1-cos(x))/sin(x))-1)**(sinh(x)/atan(x)**2), x, 0)
E
```

Derivatives are computed with the `diff` function, which recursively uses the various differentiation rules.

<sup>7</sup>The `measure` parameter of the `simplify` function lets the user specify the Python function used to determine how complex an expression is. The default `measure` function returns the total number of operations in the expression.

```
>>> diff(sin(x)*exp(x), x)
exp(x)*sin(x) + exp(x)*cos(x)
```

Integrals are calculated with the `integrate` function. SymPy implements a combination of the Risch algorithm ([Bronstein, 2005b](#)), table lookups, a reimplement of Manuel Bronstein's "Poor Man's Integrator" ([Bronstein, 2005a](#)), and an algorithm for computing integrals based on Meijer G-functions ([Roach, 1996](#); [Roach, 1997](#)). These allow SymPy to compute a wide variety of indefinite and definite integrals. The Meijer G-function algorithm and the Risch algorithm are respectively demonstrated below by the computation of

$$\int_0^{\infty} e^{-st} \log(t) dt = -\frac{\log(s) + \gamma}{s}$$

and

$$\int \frac{-2x^2(\log(x) + 1)e^{x^2} + (e^{x^2} + 1)^2}{x(e^{x^2} + 1)^2(\log(x) + 1)} dx = \log(\log(x) + 1) + \frac{1}{e^{x^2} + 1}.$$

```
>>> s, t = symbols('s t', positive=True)
>>> integrate(exp(-s*t)*log(t), (t, 0, oo)).simplify()
-(log(s) + EulerGamma)/s
>>> integrate((-2*x**2*(log(x) + 1)*exp(x**2) +
... (exp(x**2) + 1)**2)/(x*(exp(x**2) + 1)**2*(log(x) + 1)), x)
log(log(x) + 1) + 1/(exp(x**2) + 1)
```

Summations are computed with the `summation` function, which uses a combination of Gosper's algorithm ([Gosper, 1978](#)), an algorithm that uses Meijer G-functions ([Roach, 1996](#); [Roach, 1997](#)), and heuristics. Products are computed with `product` function via a suite of heuristics.

```
>>> i, n = symbols('i n')
>>> summation(2**i, (i, 0, n - 1))
2**n - 1
>>> summation(i*factorial(i), (i, 1, n))
n*factorial(n) + factorial(n) - 1
```

Series expansions are computed with the `series` function. This example computes the power series of  $\sin(x)$  around  $x = 0$  up to  $x^6$ .

```
>>> series(sin(x), x, 0, 6)
x - x**3/6 + x**5/120 + O(x**6)
```

[Section S2](#) discusses series expansions methods in more depth. Integrals, derivatives, summations, products, and limits that cannot be computed return unevaluated objects. These can also be created directly if the user chooses.

```
>>> integrate(x**x, x)
Integral(x**x, x)
>>> Sum(2**i, (i, 0, n - 1))
Sum(2**i, (i, 0, n - 1))
```



## Polynomials

SymPy implements a suite of algorithms for polynomial manipulation, which ranges from relatively simple algorithms for doing arithmetic of polynomials, to advanced methods for factoring multivariate polynomials into irreducibles, symbolically determining real and complex root isolation intervals, or computing Gröbner bases.

Polynomial manipulation is useful in its own right. Within SymPy, though, it is mostly used indirectly as a tool in other areas of the library. In fact, many mathematical problems in symbolic computing are first expressed using entities from the symbolic core, preprocessed, and then transformed into a problem in the polynomial algebra, where generic and efficient algorithms are used to solve the problem. The solutions to the original problem are subsequently recovered from the results. This is a common scheme in symbolic integration or summation algorithms.

SymPy implements dense and sparse polynomial representations.<sup>8</sup> Both are used in the univariate and multivariate cases. The dense representation is the default for univariate polynomials. For multivariate polynomials, the choice of representation is based on the application. The most common case for the sparse representation is algorithms for computing Gröbner bases (Buchberger, F4, and F5) (Buchberger, 1965; Faugère, 1999; Faugère, 2002). This is because different monomial orderings can be expressed easily in this representation. However, algorithms for computing multivariate GCDs or factorizations, at least those currently implemented in SymPy (Paprocki, 2010), are better expressed when the representation is dense. The dense multivariate representation is specifically a recursively-dense representation, where polynomials in  $K[x_0, x_1, \dots, x_n]$  are viewed as a polynomials in  $K[x_0][x_1] \dots [x_n]$ . Note that despite this, the coefficient domain  $K$ , can be a multivariate polynomial domain as well. The dense recursive representation in Python gets inefficient as the number of variables increases.

Some examples for the `sympy.polys` submodule can be found in [Section S10](#).

## Printers

SymPy has a rich collection of expression printers. By default, an interactive Python session will render the `str` form of an expression, which has been used in all the examples in this paper so far. The `str` form of an expression is valid Python and roughly matches what a user would type to enter the expression.<sup>9</sup>

```
>>> phi0 = Symbol('phi0')
>>> str(Integral(sqrt(phi0), phi0))
'Integral(sqrt(phi0), phi0)'
```

A two-dimensional (2D) textual representation of the expression can be printed with monospace fonts via `pprint`. Unicode characters are used for rendering mathematical symbols such as integral signs, square roots, and parentheses. Greek letters and subscripts in symbol names that have Unicode code points associated are also rendered automatically.

```
>>> pprint(Integral(sqrt(phi0 + 1), phi0))

$$\int \sqrt{\varphi_0 + 1} \, d(\varphi_0)$$

```

<sup>8</sup>In a dense representation, the coefficients for all terms up to the degree of each variable are stored in memory. In a sparse representation, only the nonzero coefficients are stored.

<sup>9</sup>Many Python libraries distinguish the `str` form of an object, which is meant to be human-readable, and the `repr` form, which is meant to be valid Python that recreates the object. In SymPy, `str(expr) == repr(expr)`. In other words, the string representation of an expression is designed to be compact, human-readable, and valid Python code that could be used to recreate the expression. As noted in Section 'The core', the `srepr` function prints the exact, verbose form of an expression.

Alternately, the `use_unicode=False` flag can be set, which causes the expression to be printed using only ASCII characters.

```
>>> pprint(Integral(sqrt(phi0 + 1), phi0), use_unicode=False)
 /
 |
 | -----
 | \ / phi0 + 1  d(phi0)
 |
 /
```

The function `latex` returns a  $\text{\LaTeX}$  representation of an expression.

```
>>> print(latex(Integral(sqrt(phi0 + 1), phi0)))
\int \sqrt{\phi_{0} + 1}\, d\phi_{0}
```

Users are encouraged to run the `init_printing` function at the beginning of interactive sessions, which automatically enables the best pretty printing supported by their environment. In the Jupyter Notebook or Qt Console ([Pérez & Granger, 2007](#)), the  $\text{\LaTeX}$  printer is used to render expressions using MathJax or  $\text{\LaTeX}$ , if it is installed on the system. The 2D text representation is used otherwise.

Other printers such as MathML are also available. SymPy uses an extensible printer subsystem, which allows extending any given printer, and also allows custom objects to define their printing behavior for any printer. The code generation functionality of SymPy relies on this subsystem to convert expressions into code in various target programming languages.

## Solvers

SymPy has equation solvers that can handle ordinary differential equations, recurrence relationships, Diophantine equations,<sup>10</sup> and algebraic equations. There is also rudimentary support for simple partial differential equations.

There are two functions for solving algebraic equations in SymPy: `solve` and `solveset`. `solveset` has several design changes with respect to the older `solve` function. This distinction is present in order to resolve the usability issues with the previous `solve` function API while maintaining backward compatibility with earlier versions of SymPy. `solveset` only requires essential input information from the user. The function signatures of `solve` and `solveset` are

```
solve(f, *symbols, **flags)
solveset(f, symbol, domain=S.Complexes)
```

The `domain` parameter can be any set from the `sympy.sets` module (see [Section S5](#) for details on `sympy.sets`), but is typically either `S.Complexes` (the default) or `S.Reals`; the latter causes `solveset` to only return real solutions.

An important difference between the two functions is that the output API of `solve` varies with input (sometimes returning a Python list and sometimes a Python dictionary) whereas `solveset` always returns a SymPy set object.

<sup>10</sup>See [Section S4](#) for an in depth discussion on the Diophantine submodule.

Both functions implicitly assume that expressions are equal to 0. For instance, `solveset(x - 1, x)` solves  $x - 1 = 0$  for  $x$ .

`solveset` is under active development as a planned replacement for `solve`. There are certain features which are implemented in `solve` that are not yet implemented in `solveset`, including multivariate systems, and some transcendental equations.

Some examples for `solveset` and `solve` can be found in [Section S10](#).

## Matrices

Besides being an important feature in its own right, computations on matrices with symbolic entries are important for many algorithms within SymPy. The following code shows some basic usage of the `Matrix` class.

```
>>> A = Matrix([[x, x + y], [y, x]])
>>> A
Matrix([
[x, x + y],
[y, x]])
```

SymPy matrices support common symbolic linear algebra manipulations, including matrix addition, multiplication, exponentiation, computing determinants, solving linear systems, singular values, and computing inverses using LU decomposition, LDL decomposition, Gauss-Jordan elimination, Cholesky decomposition, Moore–Penrose pseudoinverse, or adjugate matrices.

All operations are performed symbolically. For instance, eigenvalues are computed by generating the characteristic polynomial using the Berkowitz algorithm and then finding its zeros using polynomial routines.

```
>>> A.eigenvals()
{x - sqrt(y*(x + y)): 1, x + sqrt(y*(x + y)): 1}
```

Internally these matrices store the elements as Lists of Lists (LIL) ([Jones et al., 2001](#)), meaning the matrix is stored as a list of lists of entries (effectively, the input format used to create the matrix `A` above), making it a dense representation.<sup>11</sup> For storing sparse matrices, the `SparseMatrix` class can be used. Sparse matrices store their elements in Dictionary of Keys (DOK) format, meaning that the entries are stored as a dict of (row, column) pairs mapping to the elements.

SymPy also supports matrices with symbolic dimension values. `MatrixSymbol` represents a matrix with dimensions  $m \times n$ , where  $m$  and  $n$  can be symbolic. Matrix addition and multiplication, scalar operations, matrix inverse, and transpose are stored symbolically as matrix expressions.

Block matrices are also implemented in SymPy. `BlockMatrix` elements can be any matrix expression, including explicit matrices, matrix symbols, and other block matrices. All functionalities of matrix expressions are also present in `BlockMatrix`.

When symbolic matrices are combined with the assumptions submodule for logical inference, they provide powerful reasoning over invertibility, semi-definiteness,

<sup>11</sup>Similar to the polynomials submodule, dense here means that all entries are stored in memory, contrasted with a sparse representation where only nonzero entries are stored.





singularities in the middle of the interval must be specified by the user. To evaluate slowly converging limits and infinite series, mpmath automatically tries Richardson extrapolation and the Shanks transformation (Euler-Maclaurin summation can also be used) ([Bender & Orszag, 1999](#)). A function to evaluate oscillatory integrals by means of convergence acceleration is also available.

A wide array of higher mathematical functions is implemented with full support for complex values of all parameters and arguments, including complete and incomplete gamma functions, Bessel functions, orthogonal polynomials, elliptic functions and integrals, zeta and polylogarithm functions, the generalized hypergeometric function, and the Meijer G-function. The Meijer G-function instance  $G_{1,3}^{3,0}(0; \frac{1}{2}, -1, -\frac{3}{2}|x)$  is a good test case ([Toth, 2007](#)); past versions of both Maple and Mathematica produced incorrect numerical values for large  $x > 0$ . Here, mpmath automatically removes an internal singularity and compensates for cancellations (amounting to 656 bits of precision when  $x = 10,000$ ), giving correct values:

```
>>> mpmath.mp.dps = 15
>>> mpmath.meijerg([[[]],[0]], [[-0.5,-1,-1.5],[[]]], 10000)
mpf('2.4392576907199564e-94')
```

Equivalently, with SymPy's interface this function can be evaluated as:

```
>>> meijerg([[[]],[0]], [[-S(1)/2,-1,-S(3)/2],[[]]], 10000).evalf()
2.43925769071996e-94
```

Symbolic integration and summation often produce hypergeometric and Meijer G-function closed forms (see Section 'Calculus'); numerical evaluation of such special functions is a useful complement to direct numerical integration and summation.

## PHYSICS SUBMODULE

SymPy includes several submodules that allow users to solve domain specific physics problems. For example, a comprehensive physics submodule is included that is useful for solving problems in mechanics, optics, and quantum mechanics along with support for manipulating physical quantities with units.

### Classical mechanics

One of the core domains that SymPy supports is the physics of classical mechanics. This is in turn separated into two distinct components: vector algebra and mechanics.

#### *Vector algebra*

The `sympy.physics.vector` submodule provides reference frame-, time-, and space-aware vector and dyadic objects that allow for three-dimensional operations such as addition, subtraction, scalar multiplication, inner and outer products, and cross products. The vector and dyadic objects both can be written in very compact notation that make it easy to express the vectors and dyadics in terms of multiple reference frames with arbitrarily defined relative orientations. The vectors are used to specify the positions, velocities, and accelerations of points; orientations, angular velocities, and angular accelerations of

reference frames; and forces and torques. The dyadics are essentially reference frame-aware  $3 \times 3$  tensors (Tai, 1997). The vector and dyadic objects can be used for any one-, two-, or three-dimensional vector algebra, and they provide a strong framework for building physics and engineering tools.

The following Python code demonstrates how a vector is created using the orthogonal unit vectors of three reference frames that are oriented with respect to each other, and the result of expressing the vector in the  $A$  frame. The  $B$  frame is oriented with respect to the  $A$  frame using Z-X-Z Euler Angles of magnitude  $\pi$ ,  $\frac{\pi}{2}$ , and  $\frac{\pi}{3}$ , respectively, whereas the  $C$  frame is oriented with respect to the  $B$  frame through a simple rotation about the  $B$  frame's  $X$  unit vector through  $\frac{\pi}{2}$ .

```
>>> from sympy.physics.vector import ReferenceFrame
>>> A, B, C = symbols('A B C', cls=ReferenceFrame)
>>> B.orient(A, 'body', (pi, pi/3, pi/4), 'zxz')
>>> C.orient(B, 'axis', (pi/2, B.x))
>>> v = 1*A.x + 2*B.z + 3*C.y
>>> v
A.x + 2*B.z + 3*C.y
>>> v.express(A)
A.x + 5*sqrt(3)/2*A.y + 5/2*A.z
```

### **Mechanics**

The `sympy.physics.mechanics` submodule utilizes the `sympy.physics.vector` submodule to populate time-aware particle and rigid-body objects to fully describe the kinematics and kinetics of a rigid multi-body system. These objects store all of the information needed to derive the ordinary differential or differential algebraic equations that govern the motion of the system, i.e., the equations of motion. These equations of motion abide by Newton's laws of motion and can handle arbitrary kinematic constraints or complex loads. The submodule offers two automated methods for formulating the equations of motion based on Lagrangian Dynamics (Lagrange, 1811) and Kane's Method (Kane & Levinson, 1985). Lastly, there are automated linearization routines for constrained dynamical systems (Peterson, Gede & Hubbard, 2014).

### **Quantum mechanics**

The `sympy.physics.quantum` submodule has extensive capabilities to solve problems in quantum mechanics, using Python objects to represent the different mathematical objects relevant in quantum theory (Sakurai & Napolitano, 2010): states (bras and kets), operators (unitary, Hermitian, etc.), and basis sets, as well as operations on these objects such as representations, tensor products, inner products, outer products, commutators, and anticommutators. The base objects are designed in the most general way possible to enable any particular quantum system to be implemented by subclassing the base operators and defining the relevant class methods to provide system-specific logic.

Symbolic quantum operators and states may be defined, and one can perform a full range of operations with them.



```

>>> from sympy.physics.quantum import Commutator, Dagger, Operator
>>> from sympy.physics.quantum import Ket, qapply
>>> A, B, C, D = symbols('A B C D', cls=Operator)
>>> a = Ket('a')
>>> comm = Commutator(A, B)
>>> comm
[A,B]
>>> qapply(Dagger(comm*a)).doit()
-<a|*(Dagger(A)*Dagger(B) - Dagger(B)*Dagger(A))
Commutators can be expanded using common commutator identities:
>>> Commutator(C+B, A*D).expand(commutator=True)
-[A,B]*D - [A,C]*D + A*[B,D] + A*[C,D]

```

On top of this set of base objects, a number of specific quantum systems have been implemented in a fully symbolic framework. These include:

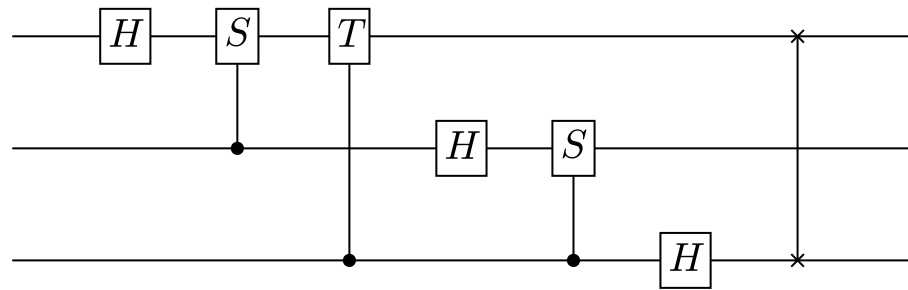
- Many of the exactly solvable quantum systems, including simple harmonic oscillator states and raising/lowering operators, infinite square well states, and 3D position and momentum operators and states.
- Second quantized formalism of non-relativistic many-body quantum mechanics ([Fetter & Walecka, 2003](#)).
- Quantum angular momentum ([Zare, 1991](#)). Spin operators and their eigenstates can be represented in any basis and for any quantum numbers. A rotation operator representing the Wigner D-matrix, which may be defined symbolically or numerically, is also implemented to rotate spin eigenstates. Functionality for coupling and uncoupling of arbitrary spin eigenstates is provided, including symbolic representations of Clebsch-Gordon coefficients and Wigner symbols.
- Quantum information and computing ([Nielsen & Chuang, 2011](#)). Multidimensional qubit states, and a full set of one- and two-qubit gates are provided and can be represented symbolically or as matrices/vectors. With these building blocks, it is possible to implement a number of basic quantum algorithms including the quantum Fourier transform, quantum error correction, quantum teleportation, Grover's algorithm, dense coding, etc. In addition, any quantum circuit may be plotted using the `circuit_plot` function ([Fig. 1](#)).

Here are a few short examples of the quantum information and computing capabilities in `sympy.physics.quantum`. Start with a simple four-qubit state and flip the second qubit from the right using a Pauli-X gate:

```

>>> from sympy.physics.quantum.qubit import Qubit
>>> from sympy.physics.quantum.gate import XGate
>>> q = Qubit('0101')
>>> q
|0101>
>>> X = XGate(1)

```



**Figure 1** The circuit diagram for a three-qubit quantum Fourier transform generated by SymPy.

```
>>> qapply(X*q)
|0111>
```

Qubit states can also be used in adjoint operations, tensor products, inner/outer products:

```
>>> Dagger(q)
<0101|
>>> ip = Dagger(q)*q
>>> ip
<0101|0101>
>>> ip.doit()
1
```

Quantum gates (unitary operators) can be applied to transform these states and then classical measurements can be performed on the results:

```
>>> from sympy.physics.quantum.qubit import measure_all
>>> from sympy.physics.quantum.gate import H, X, Y, Z
>>> c = H(0)*H(1)*Qubit('00')
>>> c
H(0)*H(1)*|00>
>>> q = qapply(c)
>>> measure_all(q)
[(|00>, 1/4), (|01>, 1/4), (|10>, 1/4), (|11>, 1/4)]
```

Lastly, the following example demonstrates creating a three-qubit quantum Fourier transform, decomposing it into one- and two-qubit gates, and then generating a circuit plot for the sequence of gates (see [Fig. 1](#)).

```
>>> from sympy.physics.quantum.qft import QFT
>>> from sympy.physics.quantum.circuitplot import circuit_plot
>>> fourier = QFT(0,3).decompose()
>>> fourier
SWAP(0,2)*H(0)*C((0),S(1))*H(1)*C((0),T(2))*C((1),S(2))*H(2)
>>> c = circuit_plot(fourier, nqubits=3)
```

## ARCHITECTURE

Software architecture is of central importance in any large software project because it establishes predictable patterns of usage and development (*Shaw & Garlan, 1996*). This section describes the essential structural components of SymPy, provides justifications for the design decisions that have been made, and gives example user-facing code as appropriate.

### The core

A computer algebra system stores mathematical expressions as data structures. For example, the mathematical expression  $x + y$  is represented as a tree with three nodes,  $+$ ,  $x$ , and  $y$ , where  $x$  and  $y$  are ordered children of  $+$ . As users manipulate mathematical expressions with traditional mathematical syntax, the CAS manipulates the underlying data structures. Symbolic computations such as integration, simplification, etc. are all functions that consume and produce expression trees.

In SymPy every symbolic expression is an instance of the class `Basic`,<sup>12</sup> the superclass of all SymPy types providing common methods to all SymPy tree-elements, such as traversals. The children of a node in the tree are held in the `args` attribute. A leaf node in the expression tree has empty `args`.

For example, consider the expression  $xy + 2$ :

```
>>> x, y = symbols('x y')
>>> expr = x*y + 2
```

By order of operations, the parent of the expression tree for `expr` is an addition. It is of type `Add`. The child nodes of `expr` are `2` and `x*y`.

```
>>> type(expr)
<class 'sympy.core.add.Add'>
>>> expr.args
(2, x*y)
```

Descending further down into the expression tree yields the full expression. For example, the next child node (given by `expr.args[0]`) is `2`. Its class is `Integer`, and it has an empty `args` tuple, indicating that it is a leaf node.

```
>>> expr.args[0]
2
>>> type(expr.args[0])
<class 'sympy.core.numbers.Integer'>
>>> expr.args[0].args
()
```

Symbols or symbolic constants, like  $e$  or  $\pi$ , are other examples of leaf nodes.

```
>>> exp(1)
E
>>> exp(1).args
```

<sup>12</sup>Some internal classes, such as those used in the polynomial submodule, do not follow this rule for efficiency reasons.

```
(
>>> x.args
(
```

<sup>13</sup>The `dotprint` function from the `sympy.printing.dot` submodule prints output to dot format, which can be rendered with Graphviz to visualize expression trees graphically.

A useful way to view an expression tree is using the `srepr` function, which returns a string representation of an expression as valid Python code<sup>13</sup> with all the nested class constructor calls to create the given expression.

```
>>> srepr(expr)
"Add(Mul(Symbol('x'), Symbol('y')), Integer(2))"
```

Every SymPy expression satisfies a key identity invariant:

```
expr.func(*expr.args) == expr
```

<sup>14</sup>`expr.func` is used instead of `type(expr)` to allow the function of an expression to be distinct from its actual Python class. In most cases the two are the same.

This means that expressions are rebuildable from their args.<sup>14</sup> Note that in SymPy the `==` operator represents exact structural equality, not mathematical equality. This allows testing if any two expressions are equal to one another as expression trees. For example, even though  $(x + 1)^2$  and  $x^2 + 2x + 1$  are equal mathematically, SymPy gives

```
>>> (x + 1)**2 == x**2 + 2*x + 1
False
```

because they are different as expression trees (the former is a `Pow` object and the latter is an `Add` object).

Another important property of SymPy expressions is that they are immutable. This simplifies the design of SymPy, and enables expression interning. It also enables expressions to be hashed, which allows expressions to be used as keys in Python dictionaries, and is used to implement caching in SymPy.

Python allows classes to override mathematical operators. The Python interpreter translates the above `x*y + 2` to, roughly, `(x.__mul__(y)).__add__(2)`. Both `x` and `y`, returned from the `symbols` function, are `Symbol` instances. The `2` in the expression is processed by Python as a literal, and is stored as Python's built in `int` type. When `2` is passed to the `__add__` method of `Symbol`, it is converted to the SymPy type `Integer(2)` before being stored in the resulting expression tree. In this way, SymPy expressions can be built in the natural way using Python operators and numeric literals.

## Extensibility

While the core of SymPy is relatively small, it has been extended to a wide variety of domains by a broad range of contributors. This is due, in part, to the fact that the same language, Python, is used both for the internal implementation and the external usage by users. All of the extensibility capabilities available to users are also utilized by SymPy itself. This eases the transition pathway from SymPy user to SymPy developer.

The typical way to create a custom SymPy object is to subclass an existing SymPy class, usually `Basic`, `Expr`, or `Function`. As it was stated before, all SymPy classes used for expression trees should be subclasses of the base class `Basic`. `Expr` is the `Basic` subclass for mathematical objects that can be added and multiplied together. The most commonly seen classes in SymPy are subclasses of `Expr`, including `Add`, `Mul`, and `Symbol`. Instances of

<sup>15</sup>See Section S3 for more information on the `sympy.logic` submodule.

Expr typically represent complex numbers, but may also include other “rings”, like matrix expressions. Not all SymPy classes are subclasses of Expr. For instance, logic expressions, such as `And(x, y)`, are subclasses of Basic but not of Expr.<sup>15</sup>

The `Function` class is a subclass of Expr which makes it easier to define mathematical functions called with arguments. This includes named functions like  $\sin(x)$  and  $\log(x)$  as well as undefined functions like  $f(x)$ . Subclasses of `Function` should define a class method `eval`, which returns an evaluated value for the function application (usually an instance of some other class, e.g., a `Number`), or `None` if for the given arguments it should not be automatically evaluated.

Many SymPy functions perform various evaluations down the expression tree. Classes define their behavior in such functions by defining a relevant `_eval_*` method. For instance, an object can indicate to the `diff` function how to take the derivative of itself by defining the `_eval_derivative(self, x)` method, which may in turn call `diff` on its args. (Subclasses of `Function` should implement the `fdiff` method instead; it returns the derivative of the function without considering the chain rule.) The most common `_eval_*` methods relate to the assumptions: `_eval_is_assumption` is used to deduce *assumption* on the object.

Listing 1 presents an example of this extensibility. It gives a stripped down version of the gamma function  $\Gamma(x)$  from SymPy. The methods defined allow it to evaluate itself on positive integer arguments, define the real assumption, allow it to be rewritten in terms of factorial (with `gamma(x).rewrite(factorial)`), and allow it to be differentiated. `self.func` is used throughout instead of referencing `gamma` explicitly so that potential subclasses of `gamma` can reuse the methods.

Listing 1: A minimal implementation of `sympy.gamma`.

```
from sympy import Function, Integer, factorial, polygamma

class gamma(Function):
    @classmethod
    def eval(cls, arg):
        if isinstance(arg, Integer) and arg.is_positive:
            return factorial(arg - 1)

    def _eval_is_real(self):
        x = self.args[0]
        # noninteger means real and not integer
        if x.is_positive or x.is_noninteger:
            return True

    def _eval_rewrite_as_factorial(self, z):
        return factorial(z - 1)

    def fdiff(self, argindex=1):
        from sympy.core.function import ArgumentIndexError
        if argindex == 1:
            return self.func(self.args[0])*polygamma(0, self.args[0])
        else:
            raise ArgumentIndexError(self, argindex)
```

The gamma function implemented in SymPy has many more capabilities than the above listing, such as evaluation at rational points and series expansion.

## Performance

Due to being written in pure Python without the use of extension modules, SymPy's performance characteristics are generally poorer than that of its commercial competitors. For many applications, the performance of SymPy, as measured by clock cycles, memory usage, and memory layout, is sufficient. However, the boundaries for when SymPy's pure Python strategy becomes insufficient are when the user requires handling of very long expressions or many small expressions. Where this boundary lies depends on the system at hand, but tends to be within the range of  $10^4$ – $10^6$  symbols for modern computers.

For this reason, a new project called SymEngine (*The SymPy Developers, 2016a*) has been started. The aim of this project is to develop a library with better performance characteristics for symbolic manipulation. SymEngine is a pure C++ library, which allows it fine-grained control over the memory layout of expressions. SymEngine has thin wrappers to other languages (Python, Ruby, Julia, etc.). Its aim is to be the fastest symbolic manipulation library. Preliminary benchmarks suggest that SymEngine performs as well as its commercial and open source competitors.

The development version of SymPy has recently started to use SymEngine as an optional backend, initially in `sympy.physics.mechanics` only. Future work will involve allowing more algorithms in SymPy to use SymEngine as a backend.

## PROJECTS THAT DEPEND ON SYMPY

There are several projects that depend on SymPy as a library for implementing a part of their functionality. A selection of these projects are listed in [Table 3](#).

## CONCLUSION AND FUTURE WORK

SymPy is a robust computer algebra system that provides a wide spectrum of features both in traditional computer algebra and in a plethora of scientific disciplines. It can be used in a first-class way with other Python projects, including the scientific Python stack.

SymPy supports a wide array of mathematical facilities. These include functions for assuming and deducing common mathematical facts, simplifying expressions, performing common calculus operations, manipulating polynomials, pretty printing expressions, solving equations, and representing symbolic matrices. Other supported facilities include discrete math, concrete math, plotting, geometry, statistics, sets, series, vectors, combinatorics, group theory, code generation, tensors, Lie algebras, cryptography, and special functions. SymPy has strong support for arbitrary precision numerics, backed by the `mpmath` package. Additionally, SymPy contains submodules targeting certain specific physics domains, such as classical mechanics and quantum mechanics. This breadth of domains has been engendered by a strong and vibrant user community. Anecdotally, many of these users chose SymPy because of its ease of access. SymPy is a dependency of many external projects across a wide spectrum of domains.

SymPy expressions are immutable trees of Python objects. Unlike many other CAS's, SymPy is designed to be used in an extensible way: both as an end-user application and as a library. SymPy uses Python both as the internal language and the user language. This

**Table 3** Selected projects that depend on SymPy.

Project name	Description
<a href="#">SymPy Gamma</a>	An open source analog of Wolfram   Alpha that uses SymPy ( <i>The SymPy Developers, 2016b</i> ). There is more information about SymPy Gamma in <a href="#">Section S11</a> .
<a href="#">Cadabra</a>	A CAS designed specifically for the resolution of problems encountered in field theory ( <i>Peeters, 2007</i> ).
<a href="#">GNU Octave Symbolic Package</a>	An implementation of a symbolic toolbox for Octave using SymPy ( <i>The Symbolic Package Developers, 2016</i> ).
<a href="#">SymPy.jl</a>	A Julia interface to SymPy, provided using PyCall ( <i>The SymPy.jl Developers, 2016</i> ).
<a href="#">Mathics</a>	A free, online CAS featuring Mathematica compatible syntax and functions ( <i>The Mathics Developers, 2016</i> ).
<a href="#">Mathpix</a>	An iOS App that detects handwritten math as input and uses SymPy Gamma to evaluate the math input and generate the relevant steps to solve the problem ( <i>Mathpix, Inc., 2016</i> ).
<a href="#">IKFast</a>	A robot kinematics compiler provided by <a href="#">OpenRAVE</a> ( <i>Diankov, 2010</i> ).
<a href="#">SageMath</a>	A free open-source mathematics software system, which builds on top of many existing open-source packages, including SymPy ( <i>The Sage Developers, 2016</i> ).
<a href="#">PyDy</a>	Multibody Dynamics with Python ( <i>Gede et al., 2013</i> ).
<a href="#">galgebra</a>	A Python package for geometric algebra (previously <code>sympy.galgebra</code> ) ( <i>Bromborsky, 2016</i> ).
<a href="#">yt</a>	A Python package for analyzing and visualizing volumetric data ( <i>Turk et al., 2011</i> ).
<a href="#">SfePy</a>	A Python package for solving partial differential equations (PDEs) in 1D, 2D, and 3D by the finite element (FE) method ( <i>Zienkiewicz, Taylor &amp; Zhu, 2013; Cimrman, 2014</i> ).
<a href="#">Quameon</a>	Quantum Monte Carlo in Python ( <i>The Quameon Developers, 2016</i> ).
<a href="#">Lcapy</a>	An experimental Python package for teaching linear circuit analysis ( <i>The Lcapy Developers, 2016</i> ).

permits users to access the same methods used by the library itself in order to extend it for their needs.

Some of the planned future work for SymPy includes work on improving code generation, improvements to the speed of SymPy using SymEngine, improving the assumptions system, and improving the solvers submodule.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

Google Summer of Code provided financial support to students who contributed to SymPy. Ondřej Čertík was supported by the Los Alamos National Laboratory. The Los Alamos National Laboratory is operated by Los Alamos National Security, LLC, for the National Nuclear Security Administration of the US Department of Energy under Contract No.



DE-AC52-06NA25396. Richard P. Muller was supported by Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. Francesco Bonazzi was supported by Deutsche Forschungsgemeinschaft (DFG) via the International Research Training Group 1524 "Self- Assembled Soft Matter Nano-Structures at Interfaces." The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Google Summer of Code.

Los Alamos National Laboratory: No. DE-AC52-06NA25396.

Sandia National Laboratories: DE-AC04-94AL85000.

International Research Training Group 1524.

### Competing Interests

Christopher P. Smith is an employee of Polar Semiconductor, Inc., Bloomington, Minnesota, United States; Mateusz Paprocki and Matthew Rocklin are employees of Continuum Analytics, Inc., Austin, Texas, United States; Andy R. Terrel is an employee of Fashion Metric, Inc, Austin, Texas, United States.

### Author Contributions

- Aaron Meurer, Christopher P. Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B. Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K. Moore, Sartaj Singh, Thilina Rathnayake, Sean Vig, Brian E. Granger, Richard P. Muller, Francesco Bonazzi, Harsh Gupta, Shivam Vats, Fredrik Johansson, Fabian Pedregosa, Matthew J. Curry, Andy R. Terrel, Štěpán Roučka, Ashutosh Saboo, Isuru Fernando, Sumith Kulal, Robert Cimrman and Anthony Scopatz wrote the paper, performed the computation work, reviewed drafts of the paper.

### Data Availability

The following information was supplied regarding data availability:

The source for the paper is at <https://github.com/sympy/sympy-paper>. The source code for SymPy is at <https://github.com/sympy/sympy>.

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.103#supplemental-information>.

## REFERENCES

- Adams WW, Loustaunau P. 1994. An introduction to Gröbner bases. Vol. 3. Boston: American Mathematical Society.

- Bailey DH, Jeyabalan K, Li XS. 2005.** A comparison of three high-precision quadrature schemes. *Experimental Mathematics* **14**(3):317–329.
- Bender CM, Orszag SA. 1999.** Advanced mathematical methods for scientists and engineers. 1st edition. Berlin Heidelberg: Springer.
- Biggs N, Lloyd EK, Wilson RJ. 1976.** Graph theory, 1736–1936. Oxford: Oxford University Press.
- Bromborsky A. 2016.** Geometric algebra/calculus modules for SymPy galgebra. Available at <https://github.com/brombo/galgebra>.
- Bronstein M. 2005a.** pmint—The Poor Man’s Integrator. Available at <http://www-sop.inria.fr/cafe/Manuel.Bronstein/pmint>.
- Bronstein M. 2005b.** Symbolic integration I: transcendental functions. New York: Springer–Verlag.
- Buchberger B. 1965.** Ein Algorithmus zum Auffinden der Basis Elemente des Restklassenrings nach einem nulldimensionalen Polynomideal. PhD thesis, University of Innsbruck, Innsbruck, Austria.
- Cervone D. 2012.** MathJax: a platform for mathematics on the Web. *Notices of the AMS* **59**(2):312–316.
- Cimrman R. 2014.** SfePy—write your own FE application. Proceedings of the 6th european conference on Python in science (EuroSciPy 2013). 65–70. Available at <http://arxiv.org/abs/1404.6391>.
- Ciurana E. 2009.** Google app engine. In: *Developing with Google App Engine. FirstPress (En ligne)*, Berkeley: Apress.
- Diankov R. 2010.** Ikfast: the robot kinematics compiler. Available at [http://openrave.org/docs/latest\\_stable/openravepy/ikfast/](http://openrave.org/docs/latest_stable/openravepy/ikfast/).
- Doorenbos RB. 1995.** Production matching for large learning systems. PhD thesis, University of Southern California.
- Faugère JC. 1999.** A new efficient algorithm for computing Gröbner bases (F4). *Journal of Pure and Applied Algebra* **139**(1–3):61–88.
- Faugère JC. 2002.** A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In: *ISSAC’02: proceedings of the 2002 international symposium on symbolic and algebraic computation*. New York: ACM Press 75–83..
- Fetter A, Walecka J. 2003.** *Quantum theory of many-particle systems*. Mineola: Dover Publications.
- Fousse L, Hanrot G, Lefèvre V, Pélissier P, Zimmermann P. 2007.** MPFR: a multiple-precision binary floating-point library with correct rounding. *ACM Transactions on Mathematical Software* **33**(2): DOI 10.1145/1236463.1236468.
- Fu H, Zhong X, Zeng Z. 2006.** Automated and readable simplification of trigonometric expressions. *Mathematical and Computer Modelling* **55**(11–12):1169–1177 DOI 10.1016/j.mcm.2006.04.002.
- Gede G, Peterson DL, Nanjangud AS, Moore JK, Hubbard M. 2013.** Constrained multi-body dynamics with Python: from symbolic equation generation to publication. In: *ASME 2013 international design engineering technical conferences and computers and*

- information in engineering conference*, New York: American Society of Mechanical Engineers, V07BT10A051–V07BT10A051.
- Goldberg D. 1991.** What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys (CSUR)* **23(1)**:5–48.
- Gosper RW. 1978.** Decision procedure for indefinite hypergeometric summation. *Proceedings of the National Academy of Sciences of the United States of America* **75(1)**:40–42.
- Gruntz D. 1996.** On computing limits in a symbolic manipulation system. PhD thesis, Swiss Federal Institute of Technology, Zürich, Switzerland.
- Horsen CV. 2015.** GMPY. Available at <https://pypi.python.org/pypi/gmpy2>.
- Hudak P. 1998.** Domain specific languages. In: Salas PH, ed. *Handbook of programming languages, vol. III: little languages and tools, chapter 3*. Indianapolis: MacMillan, 39–60.
- Hunter JD. 2007.** Matplotlib: a 2D graphics environment. *Computing in Science & Engineering* **9(3)**:90–95.
- Johansson F, The mpmath Development Team. 2014.** mpmath: a Python library for arbitrary-precision floating-point arithmetic. Version 0.19. Available at <http://mpmath.org/>.
- Jones E, Oliphant T, Peterson P, The SciPy Development Team. 2001.** SciPy: open source scientific tools for Python. Available at <http://www.scipy.org/> (accessed on 28 September 2016).
- Kane TR, Levinson DA. 1985.** Dynamics, theory and applications. New York: McGraw Hill.
- Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S, Willing C, The Jupyter Development Team. 2016.** Jupyter Notebooks—a publishing format for reproducible computational workflows. In: *Positioning and power in academic publishing: players, agents and agendas: proceedings of the 20th international conference on electronic publishing*. Amsterdam: IOS Press, 87.
- Lagrange J. 1811.** *Mécanique analytique. No. v.1*. Paris: Ve Courcier.
- Lang S. 1966.** Introduction to transcendental numbers. In: *Addison-Wesley series in mathematics*. Reading: Addison-Wesley Pub. Co.
- Lutz M. 2013.** *Learning Python*. Sebastopol: O’Reilly Media, Inc.
- Mathpix, Inc. 2016.** Mathpix — Solve and graph math using pictures. Available at <http://mathpix.com/>.
- Moses J. 1971.** Algebraic simplification: a guide for the Perplexed. In: *SYMSAC’71: proceedings of the second ACM symposium on symbolic and algebraic computation*. New York: ACM Press, 282–304 DOI [10.1145/800204.806298](https://doi.org/10.1145/800204.806298).
- Nielsen M, Chuang I. 2011.** *Quantum computation and quantum information*. New York: Cambridge University Press.
- Nijenhuis A, Wilf HS. 1978.** *Combinatorial algorithms: for computers and calculators*. second edition. New York: Academic Press.

- Oliphant TE. 2007.** Python for scientific computing. *Computing in Science & Engineering* 9(3):10–20.
- Paprocki M. 2010.** Design and implementation issues of a computer algebra system in an interpreted, dynamically typed programming language. Master's thesis, University of Technology of Wrocław, Poland.
- Peeters K. 2007.** Cadabra: a field-theory motivated symbolic computer algebra system. *Computer Physics Communications* 176(8):550–558.
- Pérez F, Granger BE. 2007.** IPython: a system for interactive scientific computing. *Computing in Science & Engineering* 9(3):21–29.
- Peterson DL, Gede G, Hubbard M. 2014.** Symbolic linearization of equations of motion of constrained multibody systems. *Multibody System Dynamics* 33(2):143–161  
DOI 10.1007/s11044-014-9436-5.
- Petkovšek M, Wilf HS, Zeilberger D. 1997.** *A = B*. Wellesley: AK Peters, Ltd., 222+. Available at <http://www.math.rutgers.edu/~zeilberg/AeqB.pdf>.
- Raymond E. 1999.** The cathedral and the bazaar. *Knowledge, Technology & Policy* 12(3):23–49.
- Roach K. 1996.** Hypergeometric function representations. In: *ISSAC'96: proceedings of the 1996 international symposium on symbolic and algebraic computation*. New York: ACM Press 301–308..
- Roach K. 1997.** In: *ISSAC'97: proceedings of the 1997 international symposium on symbolic and algebraic computation*. New York: ACM, 205–211  
DOI 10.1145/258726.2587840-89791-875-4.
- Rocklin M. 2013.** Mathematically informed linear algebra codes through term rewriting. PhD thesis, University of Chicago.
- Rocklin M, Terrel AR. 2012.** Symbolic Statistics with SymPy. *Computing in Science and Engineering* 14(3):88–93 DOI 10.1109/MCSE.2012.56.
- Rose KH. 1999.** XY-pic user's guide. Available at <http://ctan.org/tex-archive/macros/generic/diagrams/xy-pic/doc/xyguide.pdf>.
- Rosen L. 2005.** *Open source licensing: software freedom and intellectual property law*. Upper Saddle River: Prentice Hall.
- Sakurai J, Napolitano J. 2010.** *Modern quantum mechanics*. Boston: Addison-Wesley.
- Shaw M, Garlan D. 1996.** *Software architecture: perspectives on an emerging discipline*. Pittsburgh: Prentice Hall.
- Sussman GJ, Wisdom J. 2013.** *Functional differential geometry*. Cambridge: Massachusetts Institute of Technology Press.
- Tai C-T. 1997.** *Generalized vector and dyadic analysis: applied mathematics in field theory*. Vol. 9. Hoboken: Wiley-IEEE Press.
- Takahasi H, Mori M. 1974.** Double exponential formulas for numerical integration. *Publications of the Research Institute for Mathematical Sciences* 9(3):721–741.
- The Lcapy Developers. 2016.** Lcapy, a Python package for linear circuit analysis. Available at <http://lcapy.elec.canterbury.ac.nz/>.

- The Mathics Developers. 2016.** Mathics, a free, general-purpose online computer algebra system featuring Mathematica-compatible syntax and functions. Available at <https://mathics.github.io/>.
- The Quameon Developers. 2016.** Quameon, quantum Monte Carlo in Python. Available at <http://quameon.sourceforge.net/>.
- The Sage Developers. 2016.** SageMath, the sage mathematics software system. Available at <http://www.sagemath.org>.
- The Symbolic Package Developers. 2016.** The symbolic package for GNU Octave. Available at <http://octave.sourceforge.net/symbolic>.
- The SymPy Developers. 2016a.** SymEngine, a fast symbolic manipulation library, written in C++. Available at <https://github.com/symengine/symengine>.
- The SymPy Developers. 2016b.** SymPy Gamma. Available at <http://www.sympygamma.com/>.
- The SymPy.jl Developers. 2016.** SymPy.jl, a package to bring Python's SymPy functionality into Julia via PyCall. Available at <https://github.com/JuliaPy/SymPy.jl>.
- Toth VT. 2007.** Maple and Meijer's G-function: a numerical instability and a cure. Available at <http://www.vttoth.com/CMS/index.php/technical-notes/67>.
- Turk MJ, Smith BD, Oishi JS, Skory S, Skillman SW, Abel T, Norman ML. 2011.** yt: a multi-code analysis toolkit for astrophysical simulation data. *The Astrophysical Journal Supplement Series* **192**:9 DOI [10.1088/0067-0049/192/1/9](https://doi.org/10.1088/0067-0049/192/1/9).
- Zare R. 1991.** Angular momentum: understanding spatial aspects in chemistry and physics. Hoboken: Wiley.
- Zienkiewicz O, Taylor R, Zhu J. 2013.** The finite element method: its basis and fundamentals. Seventh edition. Oxford: Butterworth-Heinemann.