Word limit: 2750 word limit
Due date: Monday 31st October 2016

When supercomputers go over to the dark side

The most memorable thrillers, both in literature and film, are the ones where you get part way through the story and suddenly realise that nothing is as you thought it was. The slow realisation over the last few decades that the matter we are taught about in school accounts for only 4% of the content of the Universe – with the rest comprised of mysterious dark matter and dark energy – is as exciting and perplexing as anything from the world of fiction. After all, there is nothing more dull to a physicist than a solved problem, and they don't come much bigger than the nature of the dark stuff that accounts for over four-fifths of the matter in the Universe.

If a blockbuster is ever made about the discovery of dark matter, the next decade may well constitute the climax. New data from the Large Hadron Collider are telling us more about what dark matter can and cannot be, and the discovery of gravitational waves this year has reminded us that even hundred-year-old theories can be spectacularly confirmed in just the blink of an eye (or the chirp of a coalescing black hole). Meanwhile, direct searches for dark matter here on Earth are entering a crucial phase just as the finest-ever telescopes for searching for annihilation of dark matter in outer space start to come online. With the sensitivity of these new probes, we clearly have the technological capability to make major discoveries in the next ten years.

What is less clear, however, is how to make sense of the these disparate datasets. Imagine for a moment that you are shown a ten-acre field and told that a tiny new insect is hiding somewhere in the grass. You are also given, every few days, a new photo of a cow with a distinctive bite mark, and three distorted recordings of a strange chirrup (one of which actually turns out to just be a recording of a dying grasshopper – or a black hole merger – take your pick). After 30 years, you have seen a lot of cows, and your job is to determine the properties of the insect with no prior knowledge of it. That it is, except a vague hunch about its probable size arising from the fact that you have never seen it directly, and the fact that it hasn't managed to take down any of the cows. Now, the dark matter hunter has perhaps an easier job than the hypothetical entomologist, as she or he has a much greater abundance of data that might reveal the properties of the hunted particle, plus decades of accurate astronomical observations. However, we still have no prior knowledge of the properties of dark matter. Therefore, any one experiment that we carry out has a huge range of possible outcomes, depending on what the actual particle turns out to be. In fact, the only possible way of learning the true nature of dark matter is to piece together clues from many different experiments, and look for distinct patterns in observations that arise from totally different sources.

This puts us firmly in the realm of data science, which is enjoying an international renaissance thanks to the abundance of data provided by everything from social media applications, to government spies and melodramatic grasshoppers. The cross-pollination of techniques from a variety of disciplines is already transforming the way we look at and model the world, and research on the dark matter problem both benefits from, and contributes to, this expertise.

[Box 1:
What makes dark matter dark?
In order to qualify as “dark”, dark matter can neither emit nor absorb light.* This means that it cannot interact with photons, which are electromagnetic waves. It must therefore not feel the electromagnetic force, meaning that the dark matter particle has zero electric charge.]
*Well, not much anyway – a tiny amount is permitted in some theories, but it's as good as zero for almost all intents and purposes.

Cosmic tapas

Firm evidence for the existence of dark matter began accumulating in the 1930s, when the astronomer Fritz Zwicky noticed that the rotational motion of galaxies at the edge of the Coma galaxy cluster, which should be explained by standard Newtonian mechanics, could not be accounted for using the visible mass of the system alone. His solution was to propose the existence of a form of invisible matter, far more abundant than the visible matter, that must provide the gravitational force necessary to explain the orbital motions. The term “dark matter” arises from the fact that the matter neither emits nor absorbs light (see Box 1). By the 1980s, repeated measurements of galaxy rotation curves had shown that most galaxies were dominated by dark matter, suggesting that the Universe as a whole likely contained quite a lot more dark than visible matter. Since then, a wide variety of complementary measurements have confirmed this picture, including gravitational lensing (which relies on the bending of light by massive objects predicted by Einstein's theory of general relativity), and very precise measurements of the temperature fluctuations in the cosmic microwave background. Microwave background measurements provide information about the geometry of the Universe from the present day all the way back to the time when atoms were first formed, giving us a unique and very high precision window into its evolution and contents.

One thing unites all of these pieces of evidence for dark matter: they rely only on its gravitational interactions, telling us nothing about what the matter is actually made of. The only clues we have are that the matter should be electrically neutral (see Box 1), stable on timescales of the age of the Universe, and moving relatively slowly when galaxies and larger structures start to form. Without this final condition, galaxies and galaxy clusters would not have formed. The Standard Model of particle physics provides a compendium of every fundamental particle ever observed, and we know with certainty that none of these particles can explain the pattern of astrophysical signatures that we have collected in the past few decades.

Popular dark matter candidates include primordial black holes (black holes made fractions of a second after the big bang, not from collapsing stars), axions (ultralight particles hypothesised to solve a theoretical inconsistency in nuclear physics), and Weakly Interacting Massive Particles (WIMPs), a generic name given to any new particle heavier than about the mass of the proton that interacts with visible matter via the weak nuclear force. Even within each of these categories, there is not a unique solution to the dark matter problem, as primordial black holes could have a number of different production mechanisms (most of which require additional new physics, and are generally disfavoured both theoretically and observationally), axions come in several different types, and there are as many WIMP models as there are theoretical physicists in need of a meal ticket. If the complexity of the visible sector of the universe is anything to go by, the dark sector probably contains multiple options from this extensive menu of cosmic tapas, and the most skilled dark matter hunters must develop the ability to expertly match the tasty morsels on offer with different beverages served up by experiments.

Herding dark matter

Over the past 30 years or so, theorists have proposed and explored a huge variety of particle physics models that provide dark matter candidates, including supersymmetry, theories with extra spatial dimensions, and effective theories in which ad hoc new particles are simply added to the Standard Model without any overarching theoretical structure. All of these ideas are written in the language of quantum field theory, a mathematical formalism that is also used to provide the equations of the Standard Model. In general, we also understand how to make a generic dark matter candidate stable, by postulating one or more new fundamental symmetries that the quantum field theory must satisfy.

However, *any* attempt to extend the Standard Model leads to a huge range of potential observable consequences. This is where things start to get really interesting. Like any new arrival, almost any new particle that makes up some of the dark matter probably also comes with friends, foibles and fringe benefits. The friend particles of dark matter may behave very differently from dark matter itself. This is not all bad though – looking for dark matter's entourage and their exploits at other particle physics experiments provides another way to help pin down its identity.

Attempting to catalog all information that might in principle be useful to the dark matter problem swiftly becomes unmanageable. The Large Hadron Collider supplies a veritable torrent of data, on everything from direct searches for new, heavy particles, to suspicious decays of

Standard Model particles that might be influenced by dark matter or its friends, to Standard Model particles recoiling away from dark matter produced in collisions and trying to sneak out of CERN unseen. It also makes extremely precise measurements of the properties of the Higgs boson, which should differ slightly depending on what particles exist beyond the Standard Model. Earlier colliders provided additional constraints at lower energies, including more particle searches and precision measurements of the properties of the W and Z bosons, which, like the Higgs, can also be changed minutely by the presence of new particles. Meanwhile, fixed target experiments, in which high-energy particle beams are aimed at stationary targets, provide complementary information on, for example, dark matter candidates with some tiny electric or magnetic dipole moments. It is also possible (though not essential) that the new physics behind dark matter might disturb the neutrino sector, making measurements of neutrino properties important for a complete understanding.

Astrophysical observations useful for the particle physics of dark matter include searches for antimatter in cosmic rays, nuclear cosmic ray ratios, radio observations of distant objects, precision observations of the cosmic microwave background, and searches for high-energy neutrinos and gamma rays from space. There are also the direct searches here on Earth, which use heavily shielded, ultra-quiet apparatuses made of ultra-pure materials to try to detect individual recoils of nuclei rudely disturbed by passing dark matter. It has even been suggested that dark matter is at work in our own Sun, surreptitiously conducting heat from the core to the outer regions in such a way as to resolve tension between the standard theory of solar physics and observations from helioseismology.

To work out which theories for dark matter and its new particle friends are the mostly likely to be correct, we can performing a joint analysis of all the different searches. This essentially amounts to making up a composite likelihood function, which describes the probability that a given theory would lead to the full range of results seen at different experiments. The next steps are to choose an algorithm for scanning over the parameters of a particular theory, and to then simulate the signals expected at each experiment for each combination of parameters. The composite likelihood function then compares the predictions from all the simulations to the results of all the experiments, and gives a single number for every parameter combination that indicates how well it agrees with all data. From this, we can determine the preferred and excluded parameter ranges, and start to compare different theories. The same basic approach has already been widely applied to a variety of problems in many areas of science, including finding the correct parameters of the neutrino sector in the Standard Model and obtaining the correct interactions of the quarks.

This all sounds great in the abstract – but such global statistical fits are horribly difficult to rigorously apply to complicated theories for new particles in practice. Simulating multiple experiments carries an enormous computational burden, and so far it has always been necessary to develop approximations to the full simulations. It is also essential to correctly account for a large number of systematic uncertainties arising from imperfect experimental measurements and inaccurate (or more often, incomplete) theoretical predictions. Choosing an efficient parameter sampling algorithm is therefore paramount. For these reasons, global fits of models for new particles have traditionally only been carried out for a small number of very popular theories (mostly simple supersymmetric scenarios), using only a subset of the available data. No solution has existed for exploring generic new theories of dark matter and new particles, with generic datasets – until now.

Data mining for new particles

In 2012, our team came together to design and build the Global and Modular Beyond-Standard Model Inference Tool, or GAMBIT. Building on previous work by the particle and astroparticle communities, we aimed to design the first software package that could take *generic* theories of new physics and determine their viability by confrontation with data from all relevant particle and astroparticle experiments. Part of our strategy was to involve experts from every corner of particle and astroparticle physics, to make sure that every theoretical and experimental technique to be covered by GAMBIT would be done so carefully and rigorously. The Collaboration currently consists of approximately 30 researchers from Australia, Europe and North America, and includes members of most of the major current particle and astrophysics experiments. This allows us to tackle a much larger range of data than has ever been attempted in work of this nature.

The first challenge in designing GAMBIT was to divorce the calculation of experimental signatures from any knowledge of the fundamental parameters of the theories that might give rise to them. The idea here was to make experimental likelihood functions reusable with almost any theory. For example, given a model of the dark matter distribution in a distant galaxy, plus some new theory of particle physics, a theorist can calculate the different processes that might occur to produce gamma rays via dark matter annihilation, and then use that catalogue to obtain the expected flux of gamma rays at the Earth. A gamma ray astronomer can search for dark matter by looking for a flux of gamma rays from dark matter annihilation in a distant dwarf galaxy, and make some comment about how well the predicted flux compares with the observed one. To do this, he or she only needs to know the predicted flux – not why, how or which theories predict it. The final assessment of the viability of the theorist's model obviously relies on putting these two things together in a statistically principled way – but the key “bottleneck” of this whole calculation is the gamma-ray flux. The theory calculation can be swapped out for any other theory, so long as it predicts a gamma-ray flux. Likewise, the experimental side can be swapped out for data from any other gamma ray telescope. Similar examples crop up in essentially every sub-field of particle physics and particle astrophysics.

GAMBIT is therefore designed as a series of plug-and-play packages, where each function in the code is tagged with the quantity that it can calculate. A complex dependency resolver then stitches the functions together in the right order using techniques from graph theory, depending on which theory is to be investigated, and using which data. This allows the user to specify a theory (along with any missing calculations of derived quantities), list the experiments whose inclusion is necessary, then let the tool perform either a random scan or a statistical fit in whichever framework (Bayesian or frequentist) they desire. This system has proven remarkably versatile, allowing us to so far study supersymmetric theories, models of dark matter interacting only with the Higgs boson, and axions. The package is also readily extendible, with the addition of new models and observable calculations following a straightforward recipe.

The second fundamental challenge was how to deal with the vast array of relevant data in a way that does not “mark” it with theoretical assumptions. This needs not only fast simulation of lots of experiments, but a detailed knowledge of how to handle the related systematic and theoretical uncertainties. The conventional approach was to use simple, heuristic likelihood calculations based on published experimental results. To get the most information out of the data, and to do it as quickly as possible, we instead developed new fast simulations of the Large Hadron Collider, direct dark matter search experiments, and indirect dark matter searches with the Fermi-LAT gamma ray telescope and the IceCube neutrino telescope. Systematic uncertainties can be added as dedicated models in their own right, and scanned in tandem with the parameters of the theory for new physics.

The final major challenge to making GAMBIT work was perhaps the most obvious: speed and computational power. Over the last four years, we have had to make calculations that take hours run in less than 5 seconds, using a combination of trickery and brute force. This has meant using very fast parameter-sampling algorithms from the data science literature, and massively parallelising the code. To make efficient use of many thousands of cores at a time on clusters of multicore machines, GAMBIT uses a nested parallellisation scheme based on MPI and OpenMP. We have been fortunate enough to to receive time to run and develop it on the *Prometheus* supercomputer at Cyfronet in Krakow, the 48^{th} fastest in the world by current numbers. This has allowed us to perform real-time simulation of proton collisions at the Large Hadron Collider and simulate a range of astrophysical experiments to a high level of rigour, for hundreds of millions of different parameter combinations in various theories for new particles.

The Future

The GAMBIT code will very shortly be released as an open source public tool, along with the first physics results to use it. These will cover a range of the most popular dark matter scenarios, but the best is very much on the horizon. Many more models remain to be finally tested in full with existing data. There is plenty more exciting data to come in the near future as well. The LHC will bring down far more data at higher energies than ever before, as direct searches for dark matter ramp up to the ton scale, and indirect detection diversifies into larger telescopes with broader reaches than ever before. Axion experiments too seem set to now probe many of the most promising parts of the axion parameter space, and large-scale neutrino and rare decay searches are also due to come online in the next few years. We hope to be able to keep GAMBIT updated and ready to incorporate all these and more, and to make them available to quickly and easily test both new theories and old. At some stage soon, we might indeed track down the mysterious chirruping phantom of the Ten Acre Field – only to notice some strange marks on its wings, and glimpse from the corner of the eye something expanding rapidly off into the hundred acre field next door. By then it might be about time to ask cosmology for help again.