How Do Stars Form? Open Questions on the Stellar Initial Mass Function

Thematic Areas: ✅ Star and Planet Formation, ✅ Stars and Stellar Evolution, ✅ Resolved Stellar Populations and their Environments

Principal Author:
Name: Daniela Calzetti
Institution: Department of Astronomy, University of Massachusetts - Amherst
Email: Calzetti@astro.umass.edu
Phone: +1-413-545-2057

Co-authors: Nate Bastian (Liverpool John Moores University), Charlie Conroy (Harvard University), Bruce G. Elmegreen (IBM), John S. Gallagher (University of Wisconsin), Mark Krumholz (Australian National University), Selma E. de Mink (University of Amsterdam), Stella S. Offner (University of Texas – Austin), Elena Sabbi (STScI), Fabian R.N. Schneider (University of Heidelberg), Aida Wofford (Universidad Nacional Autónoma de México), Rosemary F.G. Wyse (The Johns Hopkins University).

Abstract:
The stars we observe are the product of natal conditions and early dynamical effects that operate on shorter timescales than the internal evolutionary times of the stars. The resulting distribution of stellar masses - the stellar ‘Initial Mass Function’ (IMF) - is a key ingredient for the interpretation of elemental abundances, galaxy evolution, and the history of stellar populations, and for predicting stellar transients (variables, explosions, gravitational waves) at all redshifts. Stars affect their surrounding environment through energy, mechanical, and chemical output; their light traces the star formation rates, star formation histories, ages, and masses of both resolved and unresolved stellar populations. These parameters heavily depend on IMF assumptions. This White Paper summarizes the progress over the past decade on observations and modeling of short timescale star formation and the IMF, assesses the scales and regimes at which IMF variations affect measurements, and discusses the challenges ahead. Open questions include the relation between cores and stars, the physical nature of the low-end characteristic mass, the existence and nature of a high-mass limit for stars, and the effect of the environment (SFR, metallicity, density,...) on stellar multiplicity and the IMF. Progress will require extensive, synergistic observations from the UV to the mm, at high angular resolution and large Field of View (FoV). Existing and upcoming optical/near/mid-IR and mm facilities must be coupled with development of UV (100-200 nm) multiplexing spectroscopic capabilities that can provide resolutions of ~3-10 pc out to ~100 Mpc, and photometric/PSF stability on arcmin FoVs.
1. Introduction

Star formation is the process that converts gas into stars and builds over time the stellar populations of galaxies. Over the past decade or so, this simple picture has acquired more and more physical foundation, and complexity. The gas in galaxies forms a continuum from atomic gas to molecular gas of progressively higher density, more compact structures, and decreasing filling factor, until it reaches the densities required for star formation (e.g., [25]). Observations with the Herschel Space Observatory have shown that stars form at the intersections of collapsing filaments [59, 5]. Stars form in a self-similar hierarchy of structures and density from dispersed complexes to compact star clusters [22, 36, 68, 69], with the spatial scale of the largest complex in disk galaxies set by shear [38]. The self-similarity of stellar complexes matches that of molecular clouds [23, 37, 51]. Models suggest that the surrounding environment both affects and is affected by the newly formed stars. The efficiency of star cluster formation, for instance, is predicted to depend on the local density of the gas [48]; observations are beginning to confirm this picture (e.g., [1]). The feedback from evolving stars, in the form of radiative, thermal, and mechanical input into the interstellar medium, produces turbulence, which is predicted to compete with gravitational instability to keep the efficiency of star formation low, at the level of a few percent on the scales of disks [52, 27, 67]. Observations with the Hubble Space Telescope have recovered the imprint of gas turbulence in the age-separation relation of star clusters [38].

For all the progress described above, key questions remain unanswered. We are missing a predictive theory of star formation that describes and connects all stages of star formation. Extant models are able to describe individual observational results, but not link them. Ambiguity among models is also a problem. For instance, debate continues on how (fragmentation, coalescence) and where massive stars form [81]. Models’ disambiguation is hampered by the complex physics involved, spanning scales over 11 orders of magnitude that cannot be dealt with even with the most powerful computers currently available.

Stars are the product both of the physical and chemical conditions of the natal gas and of the early dynamical evolution that takes place before they evolve significantly as individual systems. Most O stars, >20 $M_\odot$, are born in binary or multiple systems [73], and a significant fraction become mergers [15, 75]. Multiplicity is common at lower masses, as well [58]. Stars have widely different lifetimes (2-4 Myr for a >100 $M_\odot$ star and 10 Gyr for a 1 $M_\odot$ star) and timescales for reaching the Main Sequence (MS, from <0.1 Myr to a few 10’s of Myr for decreasing stellar mass). Massive stars’ energy and mechanical feedback could affect the growth of lower mass stars [60]. The complex interplay between natal conditions and early dynamical evolution suggests that progress towards a predictive theory requires constraints from observations of the early stages of star formation, across the full range of galactic conditions and environments. In this White Paper, the mass distribution of the products from this interplay is termed the stellar Initial Mass Function (IMF).

2. The Global Impact of the Stellar IMF

The IMF is usually described via two widely used empirical formulations: two or more connected power laws (Kroupa [50]) or a log-normal distribution (Chabrier [11], Fig. 1 left), with the original single power law (Salpeter [71]) being today a more extreme case. The low-end of the IMF impacts the masses of galaxies and traces their star formation histories (SFHs): 44% of the mass is contained in the 89% of stars with mass<1 $M_\odot$, which accumulates throughout the Hubble time, and locks a significant fraction of the baryons. Conversely, the high-end of the IMF traces any quantity related to star formation rates (SFRs), short-timescale (<500 Myr) SFHs,
large-scale feedback, ionizing photon leakage, and chemical enrichment of galaxies. Stars above \sim 10 \ M_\odot emit virtually 100\% of the ionizing photons, over 80\% of the non-ionizing UV photons, and provide the vast majority of the energy, momentum, and chemical output from stars.

**Figure 1:** (Left) The stellar IMF using the Chabrier [11, 12] representation, from 0.1 to 100 \ M_\odot. The characteristic mass, \( m_C \), is located around 0.2-0.3 \ M_\odot. The figure shows the mass range explored by studies targeting different regions in the Milky Way (MW) and other galaxies. Very Massive Stars (VMSs, out-of-frame black arrow) have masses \( >150 \ M_\odot \), beyond the plot’s range. Reproduced from [63], with permission. (Right) Top: central \sim 30 \ pc region of the nearby (3.15 Mpc) dwarf starburst galaxy NGC5253 with \( \frac{1}{2} \ Z_\odot \) (\( Z_\odot \)=solar metallicity); the magenta circles mark the location of two massive (\( >10^5 \ M_\odot \)), young (~1 Myr) star clusters embedded in dust (\( A_V \sim 2-50 \) mag). The picture is a three-color, UV-U-I, detail of a Hubble image. The two clusters, 5 and 11, coincide with the two radio peaks in the galaxy (yellow circles, [82]). Reproduced from [10], with permission. Bottom: UV spectrum of cluster 5, compared with the spectrum of VMSs in the cluster R136 in 30Dor. Specific diagnostics of VMSs [14] are: P-Cygni NV(1240) and CIV(1550) profiles, broad HeII(1640) emission, blue-shifted OV(1371) wind absorption, and an absence of SiIV(1400) P-Cygni emission/absorption. Reproduced from [78], with permission.

The ages, SFHs, and masses of galaxies and star clusters, thus, are degenerate with the IMF, whether using unresolved (e.g., via population synthesis modeling) or resolved (via color-magnitude diagrams) stellar populations. Depending on the parameter and the system measured, uncertainties can be as large as factors \sim 2-10. Presence of dust adds to the degeneracies. *JWST will obtain restframe UV/optical spectra of thousands of high-redshift galaxies that will require physical knowledge of the IMF to be interpreted.* In numerical or semi-analytical simulations of galaxies, star formation is added through empirical recipes, rather than through an a priori theory [44]. Due to lack of predictive power, models of star formation support both universal and variable IMFs [74, 39], although constraints at the low-end are being explored by leveraging the scaling properties of the characteristic mass (Fig. 1, left; [41]).

Constraints from K-band counts and dynamics suggest that a Kroupa/Chabrier IMF at the low-mass end fits well the properties of the local galaxy population (review in [8]). Recent studies using both lensing and IR spectral features also find no IMF variations in the bulk populations of early type galaxies [79, 80, 61], although galactocentric gradients, with a steeper low-end IMF towards the centers, are recovered from population synthesis models applied to sub-galactic-scale spectra [84]. Stellar counts in ultra-faint satellites of the MW out to \sim 150 \ kpc
suggest a flatter-than-Kroupa low-end IMF for decreasing metal abundance/velocity dispersion [33]. This result is, however, challenged by other studies that do not recover similar variations [91, 35], and by the potential biases induced by the limited stellar mass range probed, coupled with fitting uncertainties [64].

The number of massive stars recovered from UV observations of local star forming galaxies fully accounts for the observed ionizing photon flux, when averaged over the entire galaxy population [7]. This agrees well with the ≤5% leakage of ionizing photons from galaxies at redshift zero [77, 31]. However, systematic variations are also observed, with lower stellar and gas mass galaxies showing deficiency in their ionizing photon output [57, 54, 7], which cannot be explained with either dust obscuration or ionizing photon leakage [55]. While IMF variations have been suggested [86], the same data can be explained with stochastic IMF sampling [30, 6] and bursty star formation histories [88] in low-SFR systems.

In summary, if IMF variations exist, they appear confined to extreme (very high or very low density, e.g., [92]) environments within galaxies, and classes of galaxies (low SFR, metallicity, etc.); they do not affect the bulk properties of low-redshift galaxy populations.

3. The Challenges for the Next Decade and Beyond

3.1 Cores to Stars

Stars form in the densest gas regions, the cores. The core mass function (CMF) has a slope similar to that of the stellar IMF, and in some regions a similar but slightly higher characteristic mass [2, 47], although this seems sensitive to the definition of core [4]. The translation of the CMF to the IMF is still proving difficult, since there is no obvious physical mechanism to ensure uniform gas conversion efficiency under all conditions, or a one-to-one relation between cores and stars [64]. Simulations suggest that, unless an additional physical process such as radiative feedback prevents it, cores will simply sub-fragment as they collapse, wiping out any mapping between core and stellar masses [56, 40]. A critical question is whether the structures currently identified as cores, at masses of ~M☉ and up, remain monolithic at higher resolution, or whether they break up into complex sub-structures, indicating probable sub-fragmentation. This will likely be resolved over the next decade, as models improve and existing mm (e.g., ALMA) and upcoming near/mid-IR capabilities (JWST, WFIRST) will synergistically access the sub-pc scale required to probe cores and the earliest products of star formation (Young Stellar Objects), measure their multiplicity, disentangle the complex chemistry, and push down the mass limit to planetary masses within 2 kpc and to sub-solar levels out to ~1 Mpc.

3.2 Dynamical Evolution and the High-Mass End of the IMF

Even the initial formation of stars does not, paradoxically, fully define the IMF, because, almost immediately after formation, stars can evolve according to different paths in response to the surrounding environment. Stars that are formed in clusters do not distribute uniformly, as massive stars are found preferentially segregated toward cluster centers [42, 3, 65]. The same is found in a small fraction of MW open clusters [17]. Whether the segregation is primordial, a result of massive stars forming primarily near the centers of collapsing clouds, or dynamical, produced during an early phase of violent relaxation, is still debated [24, 53, 45].

Massive stars can result from the merging of close binaries, with an incidence of about 25% among O stars [73, 15]; effects of post-natal mass increase include apparent rejuvenation and age spreads in star clusters [20, 75]. The distribution of stellar masses that result from these ‘altered’ stars, rather than the original one, will be the one impacting the host galaxy. Whether
the result of mergers or natal conditions (e.g., rotating single stars [93, 46]), Very Massive Stars (VMSs, \( > 150 \, M_{\odot} \)) can heavily influence their surrounding environment; they can provide 25%-50% of the ionizing photon flux from the host cluster [18, 10]. However, they are extremely rare due to small number statistics and fast evolution: solid evidence for their presence has been found in 30Dor in the LMC [13], and indirect evidence in two additional galaxies [90, 78] and from detections of potential pair-instability SNe.

Robust measurements of the high-end of the IMF are still scant. Schneider+ [76] shows that 30Dor contains a 30% excess of stars in the mass range 30-200 \( M_{\odot} \) relative to the standard IMF. A few OB stars detected in LeoP (1.6 Mpc, 3% \( Z_{\odot} \)) reveal that low-\( Z_{\odot} \) massive stars are 60% fainter than high-\( Z_{\odot} \) ones [26]. Relatively isolated O stars in galaxies 1-2 Mpc away are within the reach of current spectroscopic facilities [32], and will help address high-mass star formation in low-density environments. Future AO-assisted ELTs will extend optical/nearIR spectroscopy of the atmospheres of individual stars to galaxies beyond the Local Group, to trace the \textit{time-integrated} high-end of the IMF via elemental abundance patterns [85, 16, 87].

Detection of VMSs and other massive stars in statistically large numbers requires observations of massive (>\( 10^5 \, M_{\odot} \)), young (<2 Myr) star clusters. Diagnostics of VMSs are located in the UV, where unique spectral signatures are found (Fig. 1 right, [14]). In the optical, a VMS is easily confused with Wolf-Rayet emission from a lower mass O-star. The recently approved Hubble UV Legacy Library of Young Stars (HST DDT, [62]) will obtain spectra of massive stars in the MW and Magellanic Clouds, but is unlikely to secure additional VMSs, as there are no additional massive, young star clusters beside those already surveyed. Resolving young massive clusters (sizes \( \sim 3-10 \, pc \)) in galaxies out to \( \sim 100 \, Mpc \) will provide statistics on the numbers and frequency of VMSs and other massive stars. UV detectability will be aided by the dust clearing that massive clusters perform within the first \( \sim 1-2 \, Myr \) [9, 43, 10]. As star forming regions are found throughout galaxies, multiplexing over large, \( \sim \text{arcmin} \) galactic areas will increase data collection efficiency. Thus, the characterization of VMSs and the upper end of the stellar IMF requires development of high-angular resolution, large format, sensitive multiplexing UV (100-200 nm) spectroscopic capabilities, which are not currently available.

\subsection{3.3 Stellar Multiplicity as a Function of Environment}

Knowledge of stellar multiplicity and binary frequency provides constraints to models of star formation and of the IMF [64], and of star cluster evolution [66]. Hydro-dynamical simulations show that the disks surrounding massive stars in formation tend to be unstable and break into complex systems [49, 70]. The resulting properties of the multiple systems (number of companions, distribution of separations -short vs. long periods-, distributions of mass ratios and eccentricities, and their dependence on the stellar mass) are model-dependent.

The evolution of binary systems plays a critical role in regulating the chemical enrichment of galaxies, in tracing the evolution of the Universe [21], and in the production of gravitational waves, which are targets of both LIGO and, in the future, LISA. These systems are the progenitors of SN Type Ia, X-ray binaries, double neutron stars and double black holes, and (at low metallicity) GRBs. The role of stellar multiplicity in low mass systems for planet formation is, however, ambiguous (summary in [29]).

The properties of multiple-star systems are only beginning to be quantified, and are still affected by severe observational biases, incompleteness, and limitations in the range of environments probed [19]. A recent review of binary star properties addresses selection effects in MW surveys for O-type to solar-type binaries, confirming that almost all O-stars are in
binary/multiple systems, but only half of Sun-like stars are [58]. About half of the O stars in 30Dor (LMC) are in binary systems [72].

Establishing the impact of local and global conditions (density, metallicity) on the multiplicity of B and lower-mass stars will require obtaining a census of both short period and long period binaries in massive star clusters and in galaxies other than the MW and the Magellanic Clouds. Short period (spectroscopic) binaries will be the dominion of IFUs on AO—assisted ELTs. Long period binaries require high angular resolution, and very stable PSF and photometry across the entire ~arcmin Field of View (FoV), to increase efficiency by targeting each cluster with a single pointing, and deliver proper motions as accurate as 10 μas/yr over a few years. The characterization of the members of the multiple/binary systems at all stages of evolution will require measuring the resolved massive stars’ winds and photospheric parameters, including bolometric luminosities and masses, in the UV (100-200 nm) with spatial resolution of 10’ to 100’s AU [89]. Characterizing the binary-cluster evolution cycle will require investigating very young massive clusters of different ages, in as many environments as possible, from very distributed to densely clustered.

3.4 The Outlook for the Low-Mass End

JWST, WFIRST, and AO-assisted ELTs will probe, through direct counts of low mass stars, the low-end IMF down to 0.3 M⊙ for all Milky Way dwarf satellite galaxies out to ~0.5 Mpc, and down to the hydrogen burning limit for the nearest dwarfs, thus testing theories of low mass star formation and models for the IMF characteristic mass [41]. The IR capabilities on JWST will provide a factor ~2 efficiency improvement over HST for detection of low-mass stars [34]. Measurements of proper motions will help disentangle separate populations, especially in the MW. Models of low-mass stars are steadily improving due to Kepler data and other efforts [28], and the expectation is that they will enable reliable derivations of mass functions.

The centers of early type galaxies will, however, remain unresolved, and indirect (integrated population) methods will remain the only option. The central regions of ellipticals appear to be special places in many respects; they host super supermassive BHs, present high velocity dispersions and deep potential wells, and high [a/Fe] abundance patterns; they are the centers of the first collapsing structures, with very high inferred SFRs when the bulk of the stars formed. Among these galaxies are archetypes like Centaurus A, NGC3377, M87, M60, and M49. These would be prime sites for IMF variation. The ‘cleanest’ way to model these regions is to obtain high spatial resolution (~1-2 pc) spectra to minimize contamination of MS stars by giants, in the optical regime, where M dwarfs can be separated from F and G dwarfs [83].

4. The Path Ahead

This White Paper details some of the open questions on the early stages of star formation and the stellar IMF. The challenge for the next decade is to address those questions by probing the full range of parent galaxy’s properties and environments where stars form. This will require extensive, synergistic observations from the UV to the mm, at high angular resolution and large FoV. Observations with existing and upcoming facilities at optical/near/mid-IR and mm wavelengths will need to be coupled with development of UV (100-200 nm) multiplexing spectroscopic capabilities that can provide resolutions of ~3-10 pc out to ~100 Mpc, and photometric/PSF stability on arcmin FoVs. This will need to be paired with a robust investment in theoretical and numerical modeling, informed by observations to disambiguate and link models of the different stages of star formation.
References
[19] Duchene, G., & Kraus, A. 2013, ARAA, 51, 269
[22] Elmegreen, B.G. 2010, IAU, 266, 3
[34] Geha, M. 2014, AAS, Meeting #223, id.314.03