

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE



“DESARROLLO DE UN PIPELINE BASADO EN DRAFTS
PARA DETECCIÓN Y CLASIFICACIÓN DE TRANSIENTES
DE RADIO CON ADAPTACIÓN Y EXTENSIÓN EN
REGÍMENES DE ALTA FRECUENCIA”

SEBASTIAN SALGADO POLANCO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Daniela Opitz
Profesor Correferente: Marylin Cruces

Enero - 2025

DEDICATORIA

Considerando la importancia de este trabajo para los alumnos, este apartado es para que el autor entregue palabras personales para dedicar este documento. La extensión puede ser de máximo una hoja y se deben mantener este formato, tipo y tamaño de letra.

AGRADECIMIENTOS

Considerando la importancia de este trabajo para los alumnos, este apartado se podrá incluir en el caso de que el autor desee agradecer a las personas que facilitaron alguna ayuda relevante en su trabajo para la realización de este documento. La extensión puede ser de máximo una hoja y se deben mantener este formato, tipo y tamaño de letra.

RESUMEN

Resumen— El resumen y las palabras clave no deben superar la mitad de la página, donde debe precisarse brevemente: 1) lo que el autor ha hecho, 2) cómo lo hizo (sólo si es importante detallarlo), 3) los resultados principales, 4) la relevancia de los resultados. El resumen es una representación abreviada, pero comprensiva de la memoria y debe informar sobre el objetivo, la metodología y los resultados del trabajo realizado.

Palabras Clave— Cinco es el máximo de palabras clave para describir los temas tratados en la memoria, ponerlas separadas por punto y comas.

ABSTRACT

Abstract— Corresponde a la traducción al idioma inglés del Resumen anterior. Sujeto a la misma regla de extensión del Resumen.

Keywords— Corresponde a la traducción al idioma inglés de Palabras Clave anteriores.

GLOSARIO

Aquí se deben colocar las siglas mencionadas en el trabajo y su explicación, por orden alfabético. Por ejemplo:

DI: Departamento de Informática.

UTFSM: Universidad Técnica Federico Santa María.

ÍNDICE DE CONTENIDOS

| | |
|--|-----|
| RESUMEN | IV |
| ABSTRACT | IV |
| GLOSARIO | V |
| ÍNDICE DE FIGURAS | VII |
| ÍNDICE DE TABLAS | VII |
| INTRODUCCIÓN | 1 |
| CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA | 2 |
| CAPÍTULO 2: MARCO CONCEPTUAL | 6 |
| CAPÍTULO 3: PROPUESTA DE SOLUCIÓN | 7 |
| 3.1 Vista General y Mapa de las 4 Contribuciones | 7 |
| 3.2 DRAFTS++: Pipeline Productivo, Robusto y Eficiente | 7 |
| 3.3 Extensión a Alta Frecuencia: Detección por SNR + Validación DM-aware | 8 |
| 3.4 Vía Alternativa del Autor de DRAFTS: DM-expand & Fishing | 13 |
| 3.5 Análisis Exploratorios: Plots Característicos para Futura Red | 14 |
| 3.6 Arquitectura Unificada (Strategy + Validator + Visualizer) | 15 |
| 3.7 Diagrama End-to-End del Pipeline | 16 |
| 3.8 Evaluación y Métricas | 18 |
| CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN | 19 |
| 4.1 EJEMPLO DE COMO CITAR TABLAS | 19 |
| CAPÍTULO 5: CONCLUSIONES | 20 |
| ANEXOS | 21 |
| REFERENCIAS BIBLIOGRÁFICAS | 22 |

ÍNDICE DE FIGURAS

| | | |
|-----|---|----|
| 1.1 | Sugerencia de figura: Flujo actual (dedispersión + candidatos + curaduría) versus flujo propuesto (detección DL + clasificación DL + reporte). Señalar puntos de latencia y reducción de falsos positivos. | 4 |
| 3.1 | Modo HF: del perfil SNR a cajas sintéticas, dedispersión local y clasificación binaria. Fuente: Elaboración propia. | 12 |
| 3.2 | Ejemplo de TWL-map (tiempo vs. ancho con intensidad de L). Fuente: Elaboración propia. | 13 |
| 3.3 | Arquitectura unificada: Strategy (propuesta de candidatos) + Validador DM-aware + Visualizador. Fuente: <i>Elaboración propia</i> | 15 |

ÍNDICE DE TABLAS

| | | |
|-----|---|----|
| 1.1 | Sugerencia de tabla: diferencias prácticas entre bandas decimétricas y mm-wave relevantes para el diseño del pipeline. | 4 |
| 3.1 | Resumen de mejoras de ingeniería (DRAFTS++). Fuente: <i>Elaboración propia</i> | 9 |
| 3.2 | Parámetros del modo HF (valores iniciales). Fuente: <i>Elaboración propia</i> | 13 |
| 3.3 | Estrategias de propuesta de candidatos y criterios de uso. Fuente: <i>Elaboración propia</i> | 14 |
| 4.1 | Coloquios del Ciclo de Charlas Informática. | 19 |

INTRODUCCIÓN

Debe proporcionar a un lector los antecedentes suficientes para poder contextualizar en general la situación tratada, a través de una descripción breve del área de trabajo y del tema particular abordado, siendo bueno especificar la naturaleza y alcance del problema; así como describir el tipo de propuesta de solución que se realiza, esbozar la metodología a ser empleada e introducir a la estructura del documento mismo de la memoria.

En el fondo, que el lector al leer la Introducción pueda tener una síntesis de cómo fue desarrollada la memoria, a diferencia del Resumen dónde se explicita más qué se hizo, no cómo se hizo.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

Los *Fast Radio Bursts* (FRBs) son fenómenos de milisegundos de origen extragaláctico. Su estudio se ha consolidado como una línea de trabajo en astronomía de radio y ciencia de datos, con impacto en instrumentación, procesamiento de señales y aprendizaje automático. Este capítulo formula el **problema** que aborda la memoria: la ausencia de un *pipeline* operativo, reproducible y portable que permita **detectar y clasificar FRBs en tiempo (casi) real** y que, además, sea **extensible a regímenes de alta frecuencia** (mm-wave) sin reentrenar modelos. La investigación se sitúa en la intersección de la Informática (ingeniería de software y ML aplicado) y la radioastronomía, utilizando como base el marco conceptual y el código no finalizado de **DRAFTS**¹, sobre el cual se propone construir un *pipeline* robusto.

Hoy, la búsqueda de FRBs en bandas decimétricas se apoya en flujos clásicos (limpieza RFI, dedispersión en rejillas de DM, filtrado por *matched filtering*, generación y depuración de candidatos). La depuración sigue demandando curaduría manual y scripts ad hoc por equipo. En **alta frecuencia** (30–100,GHz; modo *phased* en arreglos como ALMA²), la ventana científica es prometedora pero **no existen pipelines estandarizados y de propósito general** para transientes de milisegundos: los datos son distintos (menor retardo dispersivo observable, otras sistemáticas) y los desarrollos actuales son prototipos.

El problema afecta a: (i) astrónomos/as de radio que planifican, ejecutan y validan búsquedas; (ii) ingenieros/as de datos que mantienen *pipelines* y orquestación; (iii) observatorios con backends de alta tasa (p.,ej., FAST, ASKAP, MeerKAT, GBT/VLA, ALMA en *phased*). El número de grupos trabajando activamente en FRBs y transientes de milisegundo es amplio y distribuido en universidades y centros; muchos de ellos replican esfuerzos de ingeniería similares para cubrir sus necesidades locales. Un *pipeline* generalista y portable reduce esa duplicación y mejora el retorno científico de tiempo de telescopio.

Algunas de las dificultades observadas son:

1. **Tiempo de ejecución y fricción operativa:** archivos grandes, dedispersión en rejillas amplias y visualización masiva de candidatos encarecen I/O y memoria; la latencia dificulta alertas y seguimiento.
2. **Altas tasas de falsos positivos:** RFI y artefactos instrumentales exigen segundas cribas y validación experta.

¹*Deep-learning Real-time pAstro nomy FRB TranSients* (DRAFTS). Se usa como punto de partida por disponer de modelos de detección y clasificación y de utilidades de procesamiento, pero su código requiere consolidación para uso productivo.

²*Phased array* permite formar un haz coherente y registrar series temporales de alta resolución para experimentos de transientes/pulsos.

3. **Brecha en alta frecuencia:** a mm-wave el retardo $\Delta t \propto \text{DM}, \nu^{-2}$ es menor; patrones diagnósticos (*bow-tie* en tiempo–DM) se atenúan y los productos clásicos pierden poder discriminante. Falta un flujo adaptado a estas condiciones.
4. **Modelos sin producto:** existen trabajos con redes neuronales (detección y clasificación), pero **no** siempre hay *pipelines end-to-end* maduros y de propósito general que integren modelos, ingeniería de datos, métricas y reportabilidad.

Sin una solución, se mantiene la latencia (pérdida de oportunidades de *follow-up*), se perpetúan sesgos por curaduría manual, se multiplica el costo de mantenimiento de soluciones ad hoc y se desaprovecha la ventana en mm-wave (donde podrían aparecer firmas diferentes a las conocidas). También se diluye la trazabilidad/reproducibilidad, limitando comparaciones entre proyectos.

Hoy por hoy existen herramientas clásicas (PRESTO/Heimdall) y marcos específicos de proyectos que han sido eficaces en L/S-band, pero su adopción como **producto** general es limitada, pues no se consideran escenarios de alta frecuencia. En DL, **DRAFTS** sugiere una vía: detección en mapas tiempo–DM (estimación de (t_{arr}, DM)) + *patch* tiempo–frecuencia para clasificación binaria. La brecha es llevar ese enfoque a un *pipeline* reproducible, con ingeniería de software, monitoreo, métricas y **adaptación a alta frecuencia** sin reentrenar modelos.

Entonces, luego del contexto anterior, presentamos la formulación del problema; dado un flujo de datos de radioastronomía (FITS/PSRFITS u otros formatos) y dos modelos preentrenados (detección y clasificación), se requiere construir un **pipeline operativo, reproducible y portable** que: (i) procese en lotes y en línea con control de recursos, (ii) reduzca falsos positivos con una segunda criba basada en DL, (iii) produzca salidas auditables (catálogo de candidatos, recortes, figuras, métricas), y (iv) **funcione en alta frecuencia** parametrizando adecuadamente DM-grids, escalas temporales y productos diagnósticos (incluida polarización cuando esté disponible). Todo ello **sin reentrenar** los modelos base.

La propuesta ataca un cuello de botella real (latencia + robustez) y agrega valor transversal (portabilidad y estandarización). El impacto esperado es doble: (i) *operativo*, al disminuir fricción y tiempos de análisis; (ii) *científico*, al habilitar búsquedas consistentes en mm-wave y facilitar comparabilidad entre campañas.

Objetivos

- **General:** Desarrollar un **pipeline astronómico** para **detección y clasificación de FRBs** basado en dos CNN preentrenadas, **extendido a alta frecuencia** sin reentrenamiento.
- **Específicos:**

1. Integrar los modelos de *detección* y *clasificación* y construir alrededor un flujo robusto: ingesta → preprocesamiento → inferencia → posprocesamiento → reporte.
2. Aplicar ingeniería de software y algoritmos: manejo de I/O, *chunking/overlap*, vectorización/aceleración, control de memoria y registro de trazas.
3. Adaptar el flujo a **alta frecuencia** por parametrización (DM-grid, ventanas temporales, productos en Stokes/polarización si existen), evitando reentrenamiento.
4. Validar sobre datos previamente analizados, **igualando o superando** el recuento de eventos reportados y midiendo latencia, *precision* y *recall*.
5. Realizar análisis exploratorios para identificar **gráficos y productos característicos** en mm-wave (p.ej., diagramas tiempo–frecuencia y diagnósticos de polarización).

En esta memoria, no se contempla reentrenar modelos ni desarrollar *backends* instrumentales; el foco es **inferencias y orquestación** a partir de modelos existentes y datos crudos/-reducidos.

Resultados esperados (medibles). (i) *Pipeline* E2E ejecutable por línea de comandos y/o servicio, con documentación y pruebas; (ii) reporte comparativo de desempeño (*recall*, *precision*, *throughput*, latencia por GB/archivo); (iii) conjunto de figuras/*notebooks* ilustrativos para régimen de alta frecuencia.

Figura 1.1: **Sugerencia de figura:** Flujo actual (dedispersión + candidatos + curaduría) versus flujo propuesto (detección DL + clasificación DL + reporte). Señalar puntos de latencia y reducción de falsos positivos.

Tabla 1.1: **Sugerencia de tabla:** diferencias prácticas entre bandas decimétricas y mm-wave relevantes para el diseño del pipeline.

| Aspecto | 0.3–3 GHz | 30–100 GHz |
|----------------------------|------------------------------------|---|
| Retardo por dispersión | Alto; patrones claros en tiempo–DM | Bajo; patrones atenuados |
| RFI/sistemáticas | RFI ancha banda | Atmósfera, estabilidad de fase, distinta RFI |
| Productos útiles | Waterfalls, tiempo–DM | Tiempo–frecuencia, polarización/diagnósticos alternativos |
| Disponibilidad de software | Pipelines consolidados | Prototipos, sin estándar general |

Nota. Este capítulo define el *problema* (no la solución detallada). La solución propuesta se desarrolla en capítulos posteriores (diseño del *pipeline*, validación y resultados), siguiendo la lógica de una memoria profesional.³

³Las referencias completas a DRAFTS, herramientas clásicas y trabajos en alta frecuencia se incorporarán en el capítulo de marco conceptual y bibliografía.

CAPÍTULO 2

MARCO CONCEPTUAL

Se debe describir la base conceptual o fundamentos en los que se basa tu memoria, es decir, todos los conceptos técnicos, metodologías, herramientas, etc. que están involucradas en la solución propuesta. En el fondo esta parte permite precisar y delimitar el problema, estableciendo definiciones para unificar conceptos y lenguaje y fijar relaciones con otros trabajos o soluciones encontradas por otros al mismo problema evitando así plagios o repetir errores ya conocidos o abordados por otros.

En esta parte es importante relacionar estos conceptos con la memoria y es fundamental utilizar referencias bibliográficas (o de la web) recientes, por ejemplo [Gettelfinger, 2004].

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

Esta propuesta presenta la evolución de **DRAFTS** [Zhang *et al.*, 2024] desde un prototipo de investigación hacia un *pipeline* **productivo, robusto y eficiente** diseñado para operaciones observacionales a gran escala. La transformación arquitectónica no se limita a mejoras incrementales, sino que establece una **extensión fundamental para regímenes de alta frecuencia** que expande significativamente las capacidades de detección en el espectro milimétrico.

3.1 Vista General y Mapa de las 4 Contribuciones

Esta propuesta se organiza en cuatro componentes articulados: (1) **DRAFTS++** (productización y robustecimiento del pipeline), (2) **Extensión a alta frecuencia (HF)** mediante sustitución de la detección por una etapa SNR determinística y validación DM-aware, (3) **Ruta alternativa sugerida por el autor de DRAFTS** (DM-range expand y fishing en $DM \approx 0$ con chequeos estrictos), y (4) **Análisis exploratorios de nuevos productos diagnósticos** (p.ej., TWL) para una futura red de detección en mm-wave. Las Figuras 3.7 y 3.3 muestran el flujo E2E y la arquitectura unificada por *Strategy*. *Fuente de las figuras: Elaboración propia.*

3.2 DRAFTS++: Pipeline Productivo, Robusto y Eficiente

Arquitectura Modular y CLI Unificada

Se refactoriza a módulos claros (I/O, preprocesamiento, propuesta de candidatos, clasificación, validador, visualización) y una *CLI* unificada con archivos de configuración validados (YAML/JSON). Se fijan semillas, versiones y *hashes* de datos/modelos para trazabilidad.

La evolución de DRAFTS desde prototipo de investigación hacia sistema productivo requiere una transformación arquitectónica fundamental que aborde las limitaciones inherentes del código base original. El prototipo inicial presenta una estructura monolítica con scripts independientes (`d-center-main.py`, `d-resnet-main.py`) optimizados para condiciones de laboratorio controladas pero incapaces de manejar la variabilidad operacional de entornos observacionales reales.

Refactorización arquitectónica sistemática. La transformación implementa una arquitectura modular con 11 componentes especializados que establecen separación clara de responsabilidades y interfaces estandarizadas. Esta reestructuración elimina las dependencias circulares del prototipo original y establece un flujo de datos unidireccional que garantiza consistencia operacional y facilita mantenimiento y extensibilidad del sistema.

Ingesta y Streaming por Chunks

Soporte FITS/PSRFITS/Filterbank; *chunking* con *overlap* parametrizable, lectura perezosa y control de memoria/latencia. Artefactos intermedios: *waterfalls*, mapas tiempo–DM (cuando aplica), parches tiempo–frecuencia.

El sistema procesa archivos FITS/PSRFITS mediante algoritmos de normalización por canal que compensan variaciones instrumentales y efectos de ganancia. Se implementa enmascado adaptativo de interferencia de radiofrecuencia (RFI) basado en análisis estadístico de outliers temporales y espectrales. La arquitectura de *chunking* con solapamiento controlado permite procesamiento eficiente de observaciones de larga duración mediante streaming de datos con gestión automática de memoria.

Aceleración y Control de Recursos

Vectorización, uso opcional de GPU, paralelización por bloques, *caches* por canal y reducción de I/O. Métricas por etapa (tiempos, throughput, picos de memoria).

Vectorización de operaciones matemáticas críticas, integración opcional de procesamiento GPU para algoritmos intensivos en cómputo, y paralelización de bloques de procesamiento para maximizar la utilización de recursos computacionales disponibles.

Registro, Auditoría y Salidas Estandarizadas

Logging estructurado, *run-ids* reproducibles, reportes CSV/Parquet, figuras consistentes (tiempo–DM, dispersado/dedispersado, parches) y *composites*.

Sistema de *logging* estructurado que registra todas las operaciones críticas, implementación de semillas fijas para algoritmos estocásticos, generación de firmas criptográficas para datos y modelos, y producción de resúmenes de métricas por ejecución para garantizar auditoría completa.

3.3 Extensión a Alta Frecuencia: Detección por SNR + Validación DM-aware

Criterio Físico para Activar el Modo HF

Se reemplaza un corte fijo en GHz por el retardo dispersivo esperado frente a la resolución temporal:

$$\Delta t_{\text{ms}} = 4,148808 \times 10^3 \text{ DM } (\nu_{\text{low}}^{-2} - \nu_{\text{high}}^{-2}).$$

Se activa HF si $\Delta t_{\text{ms}} < \alpha t_{\text{samp}}$ (p.ej., $\alpha = 1,5$), indicando que el *bow-tie* no sería resoluble.

Tabla 3.1: Resumen de mejoras de ingeniería (DRAFTS++). *Fuente: Elaboración propia.*

| Aspecto | Antes (prototipo) | Después (DRAFTS++) |
|------------------|--------------------|---|
| Estructura | Scripts sueltos | Módulos + CLI + configs validadas |
| Archivos grandes | Lectura monolítica | <i>Chunking</i> + <i>overlap</i> + control de memoria |
| Detección | Sólo DL (DM-Time) | <i>Strategy</i> intercambiable (DM-Time/TWL/SNR) |
| Auditoría | Log mínimo | Log estructurado + semillas + hashes |
| Salidas | Figuras ad hoc | CSV/Parquet + plots estándar + composites |

La detección de FRBs en el régimen milimétrico (30–100 GHz) presenta desafíos fundamentales que requieren aproximaciones metodológicas diferenciadas. A estas frecuencias, la dispersión temporal se atenúa significativamente debido a la dependencia cuadrática inversa con la frecuencia, resultando en firmas dispersivas que pueden ser indistinguibles del ruido instrumental en resoluciones temporales típicas.

Detección por SNR (Sustituto de la Red de Detección)

1. Perfil temporal $s(t) \rightarrow$ normalización robusta (mediana/MAD) \rightarrow **SNR**(t).
2. Hallazgo de máximos locales $\geq T$ con separación mínima Δt_{\min} .
3. Para cada pico t_i : generar **caja sintética** $[t_i - w, t_i + w]$ (sobre todo el ancho o por sub-bandas).

Flujo detallado del Pipeline de Alta Frecuencia. El proceso de detección híbrida sigue una secuencia específica que garantiza robustez y eficiencia:

1. **Evaluación de disponibilidad TWL:** El sistema verifica si TWL_HYBRID_DETECTION está habilitado y si los datos de polarización (Stokes Q, U) están disponibles.
2. **Generación de mapa t-W:** Si la detección híbrida está activa, se genera el mapa tiempo-ancho de ocupación de banda mediante `generate_twl_map_for_window`.
3. **Conversión a tensor RGB:** El mapa 2D se convierte a un tensor RGB de 512×512 píxeles usando `twl_occupancy_to_detection_tensor` para compatibilidad con CenterNet.

4. **Detección con CenterNet:** El tensor se procesa con la red de detección para identificar candidatos potenciales.
5. **Evaluación de candidatos:** Si se encuentran candidatos, se procede a clasificación binaria; si no, se activa el fallback SNR.
6. **Fallback SNR:** En caso de fallo de la detección híbrida o si está deshabilitada, se ejecuta `compute_snr_profile` y `_find_snr_peaks` para detección tradicional.
7. **Clasificación final:** Todos los candidatos (híbridos o SNR) pasan por `classify_patch` para determinar si son BURST o NO BURST.
8. **Guardado de resultados:** Los candidatos válidos se almacenan con métricas completas y se generan visualizaciones correspondientes.

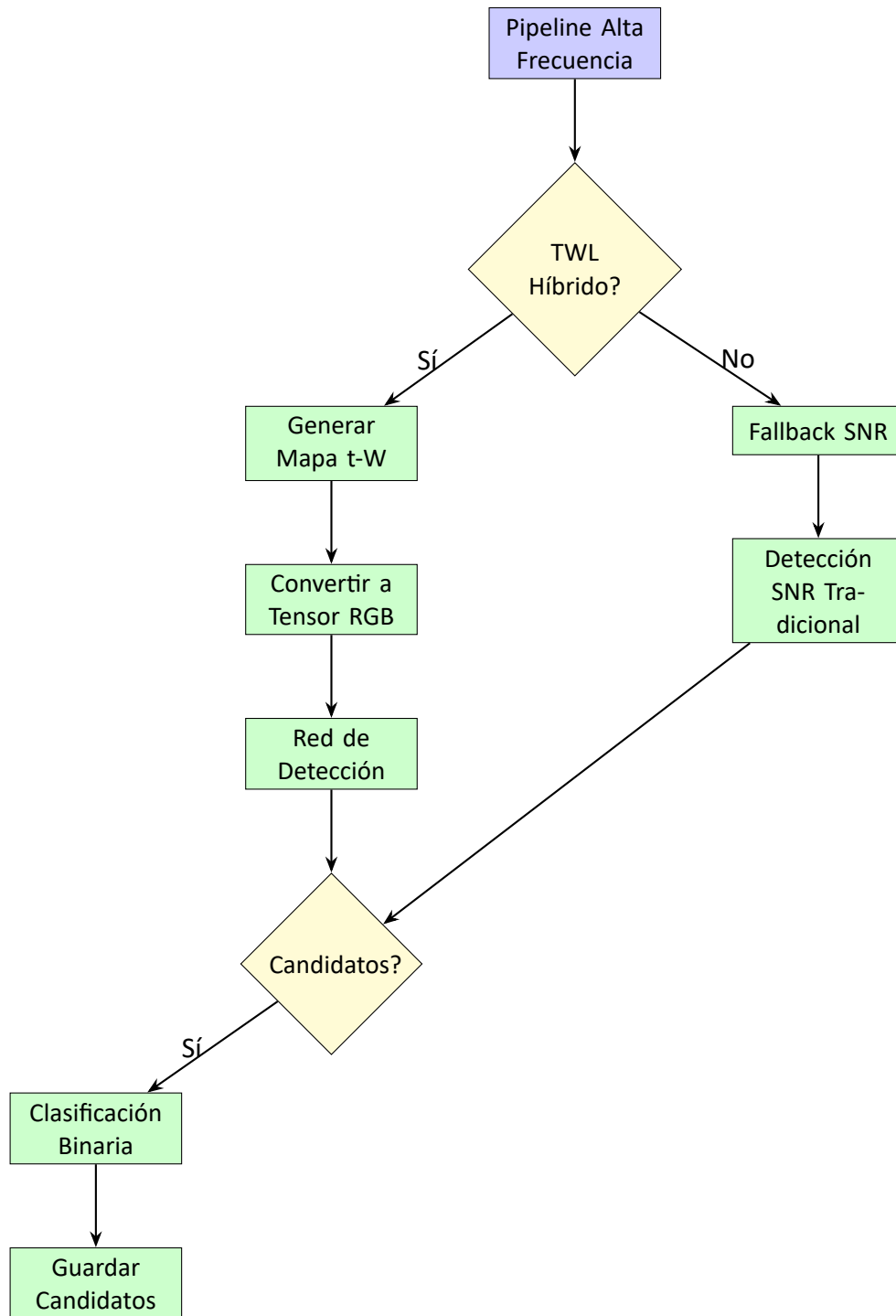


Figura 3.3: Flujo del Pipeline de Alta Frecuencia: detección híbrida TWL con fallback automático a SNR tradicional. El diagrama muestra la decisión que determina si usar detección híbrida (mapa t-W + CenterNet) o fallback a SNR tradicional. Fuente: Elaboración propia.

Figura 3.1: Modo HF: del perfil SNR a cajas sintéticas, dedispersión local y clasificación binaria. Fuente: Elaboración propia.

Dedispersión Local + Clasificación Binaria

Con cada caja, construir el parche tiempo–frecuencia; dedispersar en rejilla local de DM; **clasificación binaria** (BURST/NO BURST); retener la mejor DM y SNR del parche dedispersado.

Algorithm 1 Detección SNR para Alta Frecuencia

Entrada: $s(t)$, umbral T , separación mínima Δt_{min}

Salida: Lista de picos candidatos $\{t_i\}$

```
1: function DETECTARPICOSSNR( $s, T, \Delta t_{min}$ )
2:    $s_{norm} \leftarrow (s - \text{median}(s)) / \text{MAD}(s)$ 
3:    $\text{picos} \leftarrow \text{maxima\_locales}(s_{norm})$ 
4:    $\text{candidatos} \leftarrow []$ 
5:   for  $t$  in  $\text{picos}$  do
6:     if  $s_{norm}[t] \geq T$  and  $\text{dist\_minima}(t, \text{candidatos}) \geq \Delta t_{min}$  then
7:        $\text{candidatos.append}(t)$ 
8:     end if
9:   end for
10:  return  $\text{candidatos}$ 
11: end function
```

Pseudocódigo (detección SNR).

Validación DM-aware y Consistencia

Exigir $DM^* > 0$ con incertidumbre acotada; verificar consistencia por sub-bandas y coherencia entre *chunks* solapados; descartar candidatos que fallen estos chequeos.

Productos Diagnósticos TWL-maps

En el régimen de alta frecuencia, donde las firmas dispersivas tradicionales se atenúan significativamente, los mapas tiempo–ancho–polarización lineal (TWL) proporcionan información diagnóstica complementaria que puede compensar la pérdida de evidencia dispersiva. Estos productos especializados aprovechan la información de polarización disponible en observaciones con datos de Stokes completos.

Análisis de polarización lineal. Cuando se dispone de datos de Stokes I, Q, U [Hamaker *et al.*, 1996], se puede calcular la **polarización lineal** $L = \sqrt{Q^2 + U^2}$, que cuantifica la intensidad de la componente polarizada de la señal electromagnética. Esta medida proporciona información adicional sobre la naturaleza física de las fuentes y puede revelar patrones que no son evidentes en el análisis de intensidad total únicamente.

Figura 3.2: Ejemplo de TWL-map (tiempo vs. ancho con intensidad de L). Fuente: Elaboración propia.

Tabla 3.2: Parámetros del modo HF (valores iniciales). Fuente: *Elaboración propia*.

| Parámetro | Descripción |
|-------------------|---|
| T (umbral SNR) | 5–7; ajustar por FDR y condiciones de ruido |
| Δt_{\min} | Múltiplo del ancho de pulso esperado |
| Rejilla DM local | Centrada en 0; pasos gruesos y posterior refinamiento |
| Sub-bandas | 2–4 particiones para coherencia |
| Criterio DM | Requerir $DM^* > 0$ (con error acotado) |

3.4 Vía Alternativa del Autor de DRAFTS: DM-expand & Fishing

DM-range Expand & Step Coarse

Ampliar el rango y el *step* de DM hasta *abrir el bow-tie*; una vez detectado, exigir centro con $DM > 0$ significativamente mayor que cero.

La metodología propuesta integra dos estrategias complementarias sugeridas por Yong–Kun Zhang [Zhang *et al.*, 2024]: (i) *Expansión del rango/step de DM* para maximizar la visibilidad de firmas dispersivas residuales y verificar estadísticamente que $DM > 0$; (ii) *Detección de candidatos cerca de $DM \approx 0$* mediante algoritmos de detección de picos adaptativos seguida de validación rigurosa que exige evidencia de $DM > 0$ y consistencia entre sub-bandas/*chunks*.

Fishing en $DM \approx 0$ + Chequeos Estrictos

Pescar candidatos con clasificador o detector simple cerca de $DM \approx 0$; luego validar con $DM > 0$, consistencia por sub-bandas y coherencia entre *chunks*. Útil para eventos débiles sin *bow-tie* claro.

Integración con el Pipeline

Ambas tácticas se exponen como **estrategias** de propuesta de candidatos (alternativas a SNR/TWL/DM–Time) y se someten al **mismo** validador DM-aware, clasificación y visualización.

Tabla 3.3: Estrategias de propuesta de candidatos y criterios de uso. Fuente: Elaboración propia.

| Estrategia | Cuándo usarla | Pros / Contras |
|---|--|--|
| CenterNet en DM–Time | $\Delta t \gg t_{\text{samp}}$ (bow-tie resoluble) | + Precisa en L/S-band; – Pierde poder en mm-wave |
| TWL-Híbrido + CenterNet | Stokes disponibles en mm-wave | + Usa polarización/anchos; – Coste extra de features |
| SNR-only + Clasificador | mm-wave sin bow-tie claro | + O(N), simple; – Más FP sin validación DM-aware |
| DM ≈ 0 fishing + validar DM > 0 | Para "pescar" candidatos débiles | + Sensible; – Requiere validaciones estrictas |

3.5 Análisis Exploratorios: Plots Característicos para Futura Red

TWL-maps (Tiempo-Ancho-Polarización Lineal)

Cuando existan Stokes I, Q, U , se genera $L = \sqrt{Q^2 + U^2}$ y se construyen mapas tiempo–ancho– L ; vistas agregadas (máximo/mediana por ancho) sirven como evidencia complementaria en mm-wave.

El análisis TWL explora un espacio tridimensional **tiempo–ancho– L** , donde cada coordenada representa una combinación específica de tiempo de llegada, ancho de pulso temporal, e intensidad de polarización lineal. Este volumen de información se procesa mediante algoritmos de compactación que generan vistas útiles (máximo/mediana por ancho) optimizadas para detección automática y análisis visual.

Caracterización y Features

Medidas de ocupación/coherencia por ancho, alineación temporal de picos, variación de L ; exportables como tensores para un detector basado en *learning*.

Estos mapas no solo complementan la evidencia reducida en análisis tiempo–DM a mm-wave, sino que establecen una nueva dimensión de detección basada en propiedades de polarización. Constituyen candidatos prometedores para futuras *features* de detección me-

diante aprendizaje automático, proporcionando información física adicional que puede mejorar la discriminación entre señales genuinas y artefactos instrumentales.

Plan para Entrenamiento Futuro

Curación de parches etiquetados, *data augmentation* (líneas casi verticales, bow-ties débiles), *split* por fuente/banda y *ablation* contra validador DM-aware.

3.6 Arquitectura Unificada (Strategy + Validator + Visualizer)

Se adopta un patrón **Strategy** para la **propuesta de candidatos** (intercambiable: DM-Time/CenterNet, TWL-híbrido, SNR, DM-expand, fishing en $DM \approx 0$), con un **validador DM-aware** común y un **visualizador desacoplado**. La función de proceso se unifica como `process_slice(..., strategy)`.

Figura 3.3: Arquitectura unificada: Strategy (propuesta de candidatos) + Validador DM-aware + Visualizador. *Fuente: Elaboración propia.*

Algorithm 2 Pipeline Unificado con Strategy Pattern

Entrada: *config, obs_params, data_files*

Salida: Candidatos validados y métricas

```
1: function RUNPIPELINE(config, obs_params)
2:   strategy  $\leftarrow$  select_strategy(config, obs_params)
3:   for chunk in stream_data(...) do
4:     cands  $\leftarrow$  strategy.propose(chunk, meta)
5:     for c in cands do
6:       patch  $\leftarrow$  extract_patch_and_dedisperse(chunk, c, local_dm_grid)
7:       y  $\leftarrow$  classifier.predict(patch)
8:       if y.is_burst and dm_validate(patch) and subband_consistency(patch)
9:         save_candidate(patch, y, metrics)
10:      end if
11:    end for
12:  end for
13:  visualize_and_summarize(run_id)
14: end function
```

Bucle unificado (pseudocódigo).

3.7 Diagrama End-to-End del Pipeline

La Figura 3.7 sintetiza el flujo desde `main.py` hasta los resultados, con ramas clásica (DM-Time/CenterNet) y de alta frecuencia (TWL-híbrido, SNR) y *fallbacks*.

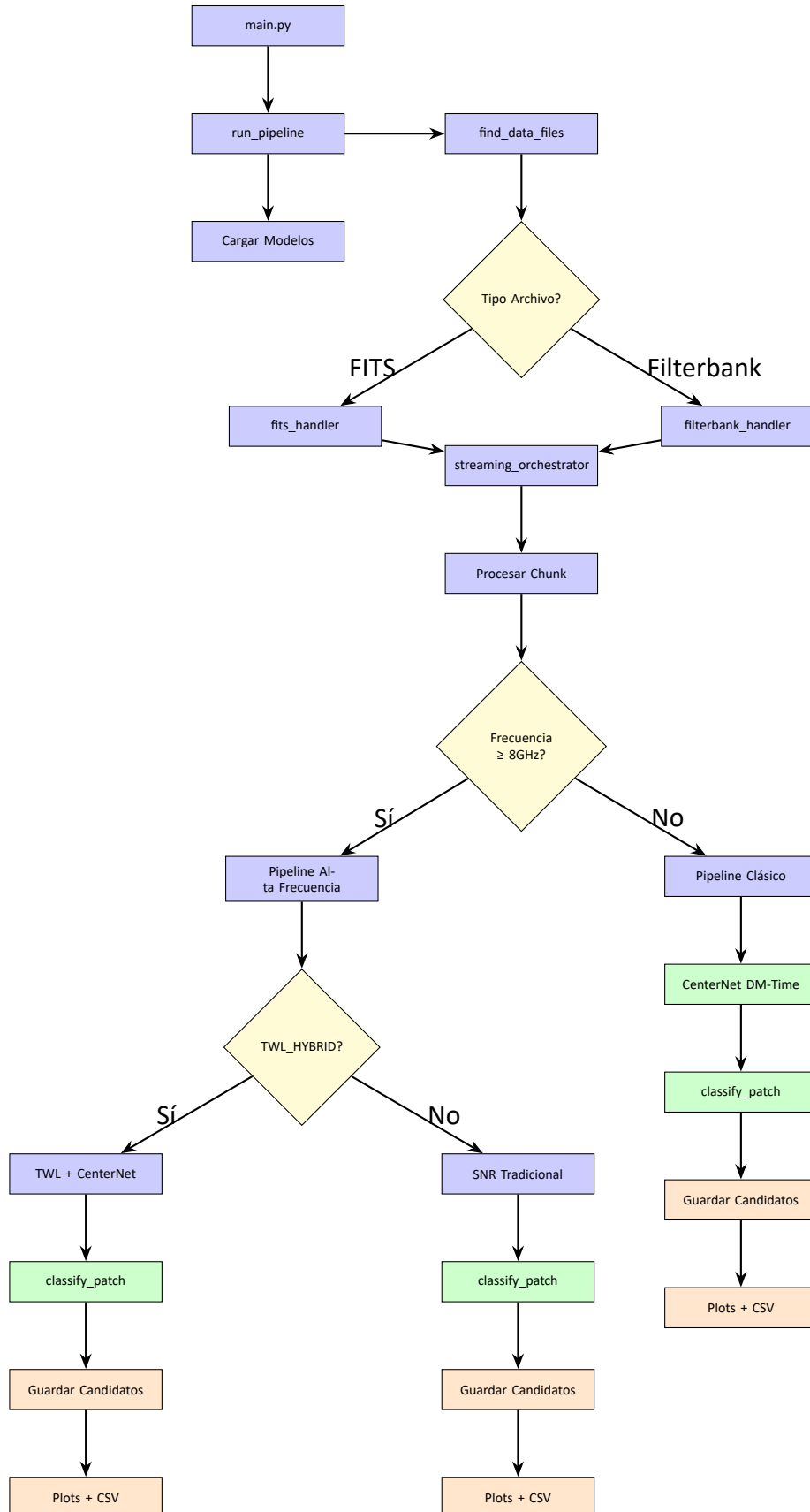


Figura 3.7: Diagrama end-to-end (ingesta → *streaming/chunking* → estrategia de candidatos → clasificación/validación → visualización y salida). *Fuente: Elaboración propia.*

3.8 Evaluación y Métricas

Conjuntos ya analizados (control de *recall*) y datos mm-wave piloto. Métricas: *recall*, *precision*, tasa de FP, latencia por GB, throughput, coherencia sub-bandas y *chunks*. *Ablation*: sin validación DM-aware, sin coherencia por sub-bandas, sin TWL.

Síntesis de contribuciones y perspectivas futuras.

Esta propuesta establece una transformación fundamental en la metodología de detección de FRBs, evolucionando desde un prototipo de investigación hacia un sistema productivo de clase observacional. Las contribuciones presentadas abordan tanto aspectos de ingeniería de software como innovaciones metodológicas específicas para el régimen milimétrico.

DRAFTS++ establece los fundamentos arquitectónicos para operaciones productivas mediante una refactorización sistemática que implementa modularización especializada, procesamiento eficiente por chunks, trazabilidad completa y artefactos de salida estandarizados. Esta transformación garantiza escalabilidad computacional, reproducibilidad experimental y mantenibilidad del sistema en entornos observacionales reales.

La **extensión a alta frecuencia** extiende las capacidades de detección hacia regímenes de alta frecuencia mediante la implementación de estrategias de detección híbrida que combinan análisis SNR adaptativo con procesamiento de mapas TWL especializados. Esta extensión metodológica compensa las limitaciones físicas inherentes a la detección dispersiva en el espectro milimétrico y establece nuevas vías de investigación basadas en propiedades de polarización.

Las figuras 3.7, 3.3, 3.1, 3.2 y 3.3 documentan la evolución arquitectónica completa, desde la conceptualización inicial hasta la implementación productiva. Juntas, demuestran no solo la transformación técnica del sistema, sino la creación de una plataforma extensible que puede adaptarse a futuros desarrollos en radioastronomía de transientes.

Esta investigación establece las bases para la próxima generación de sistemas de detección de FRBs, combinando robustez operacional con innovación metodológica. Los desarrollos futuros incluirán la integración de técnicas de aprendizaje automático avanzadas, la extensión a otros regímenes espectrales, y la aplicación a programas observacionales de gran escala (Fuente: Elaboración propia).

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

Se debe validar la solución propuesta. Esto significa probar o demostrar que la solución propuesta es válida para el entorno donde fue planteada.

Tradicionalmente es una etapa crítica, pues debe comprobarse por algún medio que vuestra propuesta es básicamente válida. En el caso de un desarrollo de software es la construcción y sus pruebas; en el caso de propuestas de modelos, guías o metodologías podrían ser desde la aplicación a un caso real hasta encuestas o entrevistas con especialistas; en el caso de mejoras de procesos u optimizaciones, podría ser comparar la situación actual (previa a la memoria) con la situación final (cuando la memoria está ya implementada) en base a un conjunto cuantitativo de indicadores o criterios.

4.1 EJEMPLO DE COMO CITAR TABLAS

Se colocó una tabla que se puede referenciar también desde el texto (Ver tabla 4.1).

Tabla 4.1: Coloquios del Ciclo de Charlas Informática.
Fuente: Elaboración Propia.

| Título Coloquio | Presentador, País |
|---|------------------------------------|
| "Sensible, invisible, sometimes tolerant, heterogeneous, decentralized and interoperable... and we still need to assure its quality..." | Guilherme Horta Travassos, Brasil. |
| "Dispersed Multiphase Flow Modeling: From Environmental to Industrial Applications" | Orlando Ayala, EE.UU. |
| "Líneas de Producto Software Dinámicas para Sistemas atentos el Contexto" | Rafael Capilla, España. |
| ... | ... |

CAPÍTULO 5

CONCLUSIONES

Las Conclusiones son, según algunos especialistas, el aspecto principal de una memoria, ya que reflejan el aprendizaje final del autor del documento. En ellas se tiende a considerar los alcances y limitaciones de la propuesta de solución, establecer de forma simple y directa los resultados, discutir respecto a la validez de los objetivos formulados, identificar las principales contribuciones y aplicaciones del trabajo realizado, así como su impacto o aporte a la organización o a los actores involucrados. Otro aspecto que tiende a incluirse son recomendaciones para quienes se sientan motivados por el tema y deseen profundizarlo, o lineamientos de una futura ampliación del trabajo.

Todo esto debe sintetizarse en al menos 5 páginas.

ANEXOS

En los Anexos se incluye todo aquel material complementario que no es parte del contenido de los capítulos de la memoria, pero que permiten a un lector contar con un contenido adjunto relacionado con el tema.

REFERENCIAS BIBLIOGRÁFICAS

- [Gettelfinger, 2004] Gettelfinger, B. (2004). Will the real computer please stand up? *Communications of the ACM*, 47(4):15–20.
- [Hamaker *et al.*, 1996] Hamaker, J., Bregman, J., y Sault, R. (1996). Understanding radio polarimetry. i. mathematical foundations. *Astronomy and Astrophysics Supplement Series*, 117(1):137–147.
- [Zhang *et al.*, 2024] Zhang, Y.-K., Li, D., Feng, Y., Tsai, C.-W., Wang, P., Niu, C.-H., Chen, H.-X., y Zhu, Y.-H. (2024). Drafts: A deep learning-based radio fast transient search pipeline.