15TH IEA HEAT PUMP CONFERENCE
MAY 26–29, 2026 | HOFBURG
VIENNA 2026
DECARBONISATION THROUGH INNOVATION

Why does Control of Heat Pump Cascade matter?

Insights for Monovalent Air-Source Systems in Large, Non-Retrofitted Multifamily Buildings

Omar Montero[a], Pauline Brischoux[a], Carolina Fraga[b], Matthias Rüetschi[b], Edouard Vionnet[b], Nicole Calame[c], Fabrice Rognon[c], Martin Patel[a], and Pierre Hollmuller[a]

[a]Energy Systems Group, University of Geneva, 66 Bd Carl Vogt, 1211 Genève, Switzerland
[b]Services Industriels de Genève, Chemin du Château-Bloch 2, 1219 Le Lignon, Genève, Switzerland
[c]CSD Ingénieurs SA, Avenue des Sports 14, 1400 Yverdon-les-Bains, Switzerland

Abstract

This study investigates the operation of multi-stage air-source heat pumps (AHSPs) in non-retrofitted multifamily buildings (MFBs) with high ratios of space heating (SH) to domestic hot water (DHW) demand and limited storage. Hourly monitoring over one year in an eight-story building in Geneva reveals that compressor activation delay significantly influences cycle duration and frequency. Complementary dynamic simulations evaluate step-by-step activation of dedicated SH and DHW HP units, under varying outdoor temperatures, storage volumes, and HP sizing. Results show that adjustable activation delays are essential to balance cycle duration, thermal comfort, and equipment longevity, particularly during cold periods. Sensitivity analysis highlights the importance of storage volume for mitigating short cycling and maintaining SH/DHW temperature setpoints, whereas oversizing HPs alone is insufficient. Findings provide guidance for cascading strategies in large MFBs and emphasize the need for integrated HP control to optimize efficiency, comfort, and system reliability.

Selection and/or peer-review under the responsibility of the organizers of the 15th IEA Heat Pump Conference 2026.

Keywords: air-source heat pump; multifamily building; cascade; comfort; in-situ montoring; numerical simulation

1. Introduction

Heat pumps (HPs) are increasingly deployed for space heating (SH) and domestic hot water (DHW) in multifamily buildings (MFB). In non-retrofitted buildings, integration is challenging due to limited space in the boiler room and buffer storage, high SH demand relative to DHW, and legacy hydraulic layouts. These constraints create a trade-off when implementing sequential activation (cascading) of multiple HP units or compressors: delaying successive stages can lengthen operating cycles, thereby reducing short-cycling and compressor wear, but may transiently compromise temperature setpoints.

Sequential activation improves part-load efficiency and system reliability by engaging only the necessary compressors and avoiding unnecessary starts. Laboratory and simulation studies have shown that modulation or controlled staging reduces the number of starts, lowers energy losses, and extends compressor lifetime [1–3]. However, staged operation introduces a dilemma: enforcing minimum run-times or stage delays reduces wear but can delay capacity response, potentially affecting thermal comfort and DHW delivery [2,4]. Most prior work focuses on new or lab-test buildings, leaving a knowledge gap for existing MFB with limited thermal storage and high SH to DHW ratios.

Long-term field evidence remains limited. Only a few studies report hourly operation over one year in non-retrofitted MFBs [5,6]. These datasets confirm that small storage exacerbates short cycling and temperature fluctuations when SH dominates DHW and highlights the critical role of control strategies and buffering in maintaining temperature setpoints while protecting equipment.

Despite these insights, important gaps remain: (i) quantitative evidence of how stage-delay settings affect cycling behavior and SH/DHW temperature control in real MFBs, and (ii) dynamic simulation of stage-delay settings in existing MFBs with high SH-to-DHW load ratios and limited storage.

The present study addresses these gaps by combining one-year, hourly in-situ monitoring with dynamic simulation of multi-stage HP operation. It focuses on the trade-off between delaying the activation of successive stages to achieve longer cycles and maintaining temperature setpoints under variable loads.

2. Monitoring

2.1. System description

The analyzed P&D project (Figure 1) concerns a non-retrofitted MFB constructed in 1972, with a heated floor area of 4,047 m², which has not undergone thermal refurbishment. The measured annual heat demand amounts to 94 $kWh_{th}/m^2 \cdot yr$, with 68% for SH and 32% for DHW. SH is supplied by radiators with supply temperatures of 50°C at 0°C outdoor temperature. Objective of the P&D project was to show that switch from fossil boiler to HP in non-refurbished MFB is feasible, with substantive CO_2 savings. The installation of a ground-source HP is not allowed by local regulations; therefore, the existing gas boiler was replaced by two large industrial ASHP (2 × 156 kW at 7/35 °C) providing both SH and DHW. Due to limited space in the boiler room, the storage volume is restricted to 1 m³ for SH and 2 m³ for DHW. Monitoring over two years of operation shows an SPF of 2.5 and significant CO_2 savings in [7], while the specific issue of HP cascading is not addressed.

Figure 1. Non-retrofitted MFB (left) equipped with two air-source HP on the roof (right).

Meeting of the required load is achieved by step-by-step activation of the total HP capacity, which consists in sequential activation of the HP units and the compressors within each unit (4 compressors per HP).

Initially, cascading was coordinated by the HP's internal logic, based on the return temperature from the boiler room. This resulted in continuous operation of the circulation pump, even when no heat production was required, causing electric consumption without heat production and strongly deteriorating the coefficient of performance at the system level.

In a second step, this issue was resolved by installing an external MCR (Master Control Relay), for sequential activation of the HP units. However, since the MCR did not directly control the individual compressors, the latter continued to be activated by each HP's internal logic, via the programmed activation delay (fixed time interval before the next compressor starts, if demand is unmet): a short delay enables a quick response to load, whereas a longer delay slows compressor activation and extends cycle duration.

2.2. Monitoring results

The monitoring data displayed in Figure 2 corresponds to the optimized period, with MCR control of the HP units, during which two delays for activation of the HP compressors were tested: 10 min or 0 min (all four compressors started simultaneously). It shows the average cycle duration and the daily number of cycles as a function of outdoor temperature.

For outdoor temperatures below 20 °C, where both SH and DHW may be required, system behavior is highly sensitive to the delay setting. With 0 min, the temperature setpoint is quickly reached but cycles become very short (< 20 min) and frequent (up to 50 cycles/day), increasing the risk of compressor wear. This 20 min limit is consistent with professional guidelines, which generally restrict heat pump operation to approximately

3 cycles per hour. With 10 min delay, the daily number of cycles remains below 20, but their duration extends significantly, with a median of 120 min and occasional values above 4 h, which could impair thermal comfort.

For outdoor temperatures above 20 °C, corresponding to DHW cycles only, the influence of the activation delay is less pronounced. Reducing the delay from 10 to 0 min mainly shortens the median cycle duration from 53 to 35 min, while the number of daily starts remains almost unchanged (around 6 cycles per day).

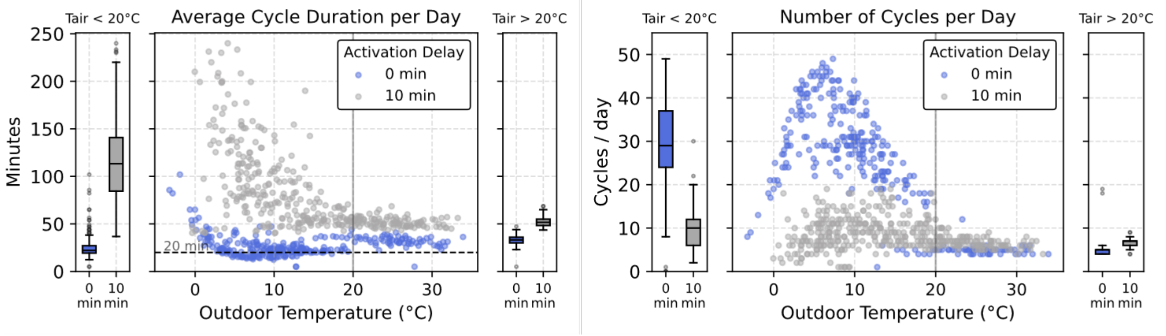

Figure 2. Average cycle duration (left) and number of cycles (right) with compressor activation delays of 0 and 10 minutes, as a function of outdoor temperature (daily values).

3. Simulation

To further analyze the seasonal behavior observed in the monitoring campaign, dynamic simulations are performed in TRNSYS (1-min time step, full-year run). The model was validated over a one-year period, using the hydraulic layout, components sizing and control strategy observed on site [7].

As a notable difference with the P&D project, for which SH and DHW are produced alternatively by the same set of HPs (with priority to DHW), the simulated system relates to separate HP units for both services (service-dedicated HPs). The reasons for such a choice are two-fold. Firstly, it allows independent observation of system behavior for each service. Secondly, if space is not an issue, this type of system has the advantage of: i) lower installed HP capacity, since SH can be delivered 24/24h; ii) less complex system hydraulics and control; iii) possibility to use refrigerants adapted to the temperature level (not analyzed in this study).

Furthermore (for both SH and DHW), the total HP capacity is subdivided into 4 identical stages, which are activated successively, with a user-defined delay (Δt) which can be varied from 0 to 18 min (fixed value over the year), in 3-min increments. As shown in Figure 3, instantaneous activation of the full HP capacity (Δt = 0 min) induces shorter cycles than its step-by-step activation.

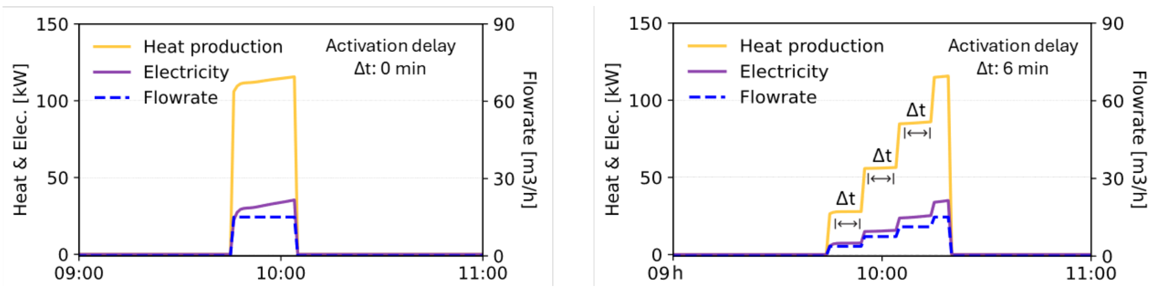

Figure 3. Activation of the HP capacity in instantaneous mode (left) versus step-by-step mode (right).

3.1. Reference case

In the first step (reference case), heat load and corresponding system sizing are set as follows:

- **Heat demand**: SH is derived from the energy signature of the pilot project (in hourly values), adjusted to the reference climate to reach an annual demand of 100 kWh/m^2, which corresponds to the median SH demand of Geneva's postwar MFB stock. The DHW demand is set to 35 L/person·day at 60 °C for 114 occupants and modelled using a prescribed hourly draw-off profile defined by Swiss normative values, with a cold water input temperature of 10 °C. Including 30% storage and distribution losses, this results in an annual DHW demand of 30 kWh/m^2·yr, consistent with observed values for multifamily buildings in Geneva.

- **HP capacity** is calculated for covering the maximal hourly load of each service: 208 kW for SH, 88 kW for DHW (both capacities given at 7°C/45°C operation).
- **Storage volume** is set to 3.3 m³ for SH (15.8 L/kW), sized to guarantee a minimal 20-minute HP cycle at 15°C outdoor, when the SH load is at its lowest and effective HP capacity at its highest. For DHW, the storage volume is set to 1.9 m³ (21.5 L/kW), to guarantee a maximum 1-hour cycle at −5/60 °C, to limit comfort issues when the effective HP capacity is at its lowest.

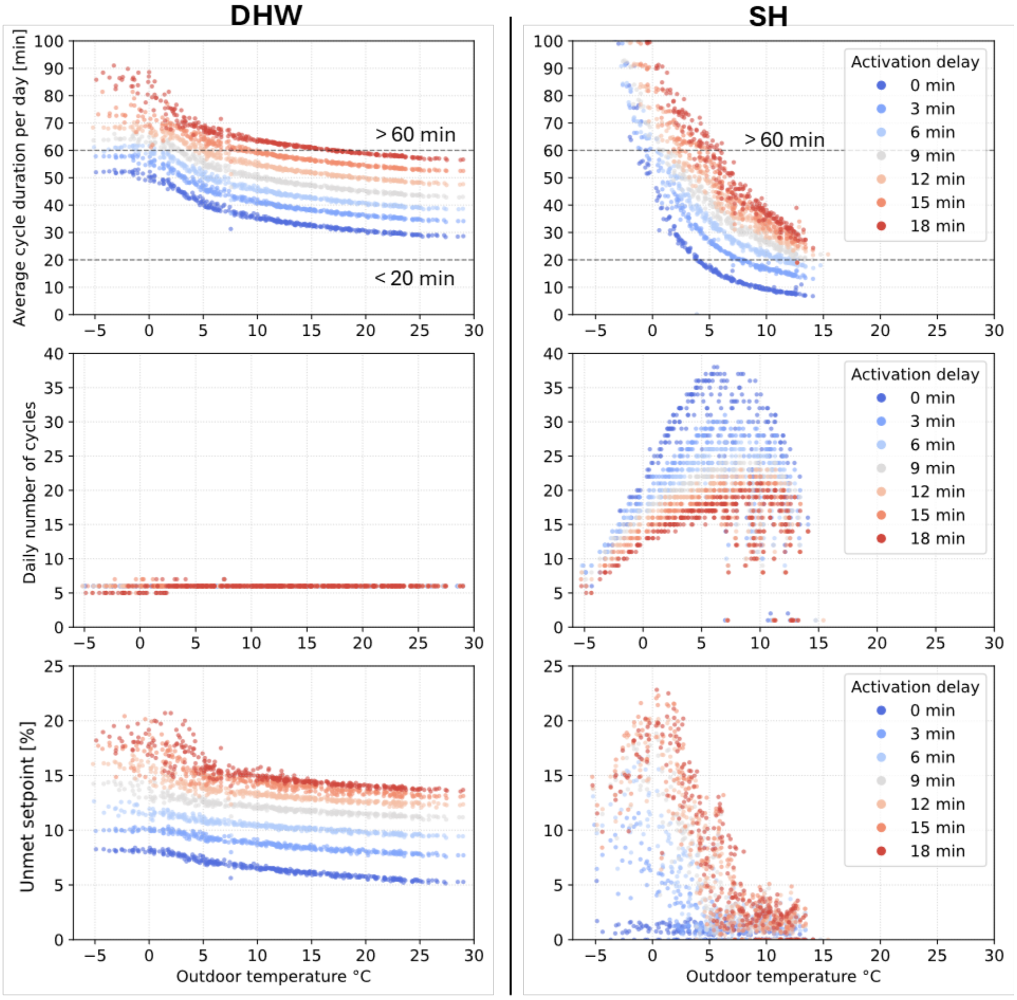

Figure 4. Daily performance indicators as a function of outdoor temperature.

Figure 4 reports daily HP performance in relation to outdoor temperature, for DHW (left) and SH (right), in terms of average cycle duration, the daily number of cycles, as well as frequency of unmet temperature setpoint. Latter is defined as follows: i) for DHW it is given by the number of hours (relative to 24 h), during which the temperature at the top of the storage tank is lower than 50°C (value recommended in Swiss standards to reduce the risk of legionella development); ii) for SH it is defined as the number of hours (relative to 24 h), during which the top of the storage tank drops 5K below the setpoint of the heating curve. Note that the 5K tolerance corresponds to a drop in distribution temperature which we suppose could be temporarily absorbed by the building inertia.

For DHW, the daily number of cycles is essentially stable throughout the year (6 cycles per day) because the DHW draw-off pattern is constant. However, the cycles become somewhat longer when the outdoor temperature drops below 5°C, due to a related drop of COP and effective HP capacity (larger ΔT between heat demand and source). More importantly, the cycle duration is directly related to the activation delay: at −5 °C, the cycle duration stays below 60 min for activation delays < 6 min but rises to ≈ 90 min with a 18 min delay.

Consequently, the desired 50°C setpoint is unmet each day for a few hours, with a frequency which increases with the activation delay, as well as with the drop in outdoor temperature. This is, however, linked to a strict definition of DHW availability at 50°C, which is usually not necessary for direct use and should in principle not pose health problem if the temperature rises above this setpoint within 24 h (which is warranted, since the daily indicator never rises above 20%).

For SH, the cycle behavior is even more sensitive to the outdoor temperature. When latter drops the building load rises and the effective HP capacity decreases, so the HP needs more runtime to reach the SH setpoint, so that cycles become longer. Hence, early in the heating season the daily number of cycles increases as the tank discharges faster. Below 5°C, successive HP activations grow long enough to overlap, leaving no off-time and effectively collapsing multiple starts into a single continuous run; the daily cycle number hence falls, reproducing the bell-shaped pattern seen in the monitoring (Figure 2). Cycles are also very sensitive to the activation delay: above 5°C, delays below 6 min induce very short cycles (< 20 min); below 0°C delays above 3 min induce very long cycles (> 60 min).

Consequently, as for DHW, the frequency of unmet SH temperature setpoints increases both with the activation delay and the drop in outdoor temperature. While a 0 min delay results in almost 0% unmet setpoint hours over the entire heating season, more significant setpoint deviations occur with higher delays, reaching up to 20% of the day for the 18 min delay.

3.2. Sensitivity analysis

As a complement, we perform a sensitivity analysis concerning sizing of following SH elements, which may impact both cycle behavior and the frequency of unmet SH temperature setpoints: i) larger or smaller storage volume (6 m^3 or 1 m^3, instead of 3.3 m^3), which reflect the range of space availability observed in real boiler rooms; ii) over- or under-sizing of HP capacity (236 or 175 kW, instead of 208 kW), reflecting typical uncertainties where heating demand is often estimated rather than measured and may lead to sizing deviations. The boxplots of Figure 5 represent the distribution of the daily performance indicators over the entire heating season, for three different activation delays (0, 9 or 18 min).

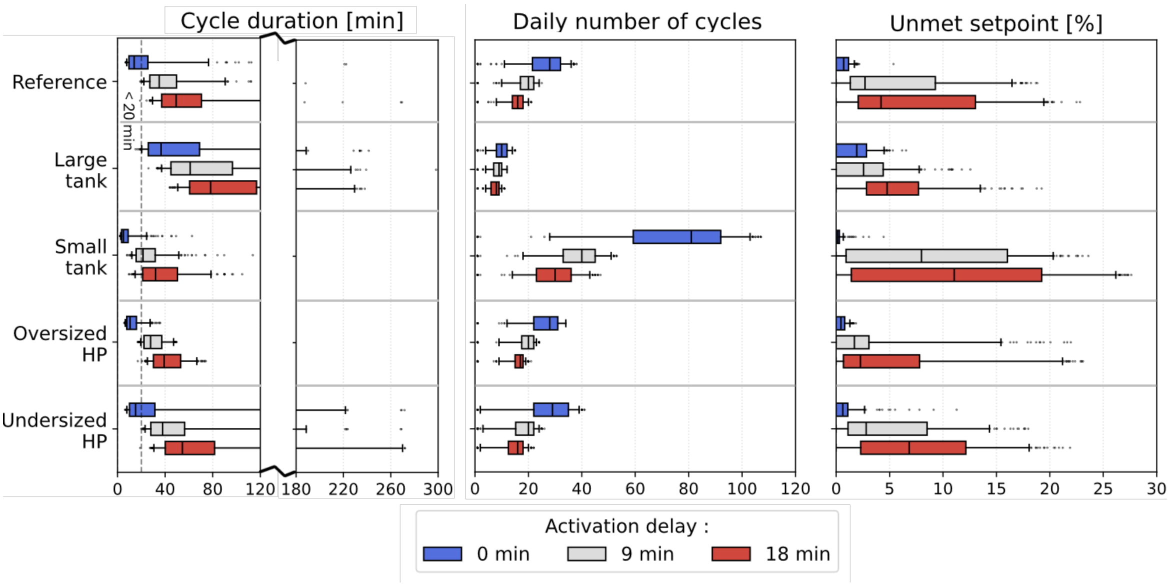

Figure 5. Distribution of daily SH performance indicators over the heating season (median, quartiles, deciles and outliers).

These synthetic results can be summarized as follows:

- **Storage tank**: A larger storage tank tends to increase the cycle duration, which barely falls below the 20 min threshold anymore and to reduce the frequency of unmet SH temperature setpoints. On the contrary, a smaller storage tank induces very short cycles and a higher risk of unmet SH temperature setpoints.
- **HP capacity**: With an oversized HP, the frequency of unmet temperature setpoint decreases, but on the cost of shorter cycles. On the contrary, with an undersized HP, the cycle duration increases, particularly during peak load (> 2 h), so does the median frequency of unmet temperature setpoint, despite maximum values remaining alike the ones of the reference case.

4. Discussion and Conclusion

Simulation of HP system for a typical non-retrofitted MFB in Geneva (with SH three times higher than DHW), with separate units for SH and DHW production, yields following insights.

For SH, a clear trade-off emerges between limiting HP wear and maintaining temperature setpoints. Above 5 °C outdoor temperature, achieving cycle durations longer than 20 min requires a step-by-step activation of HP capacity. In contrast, maintaining the SH setpoint, especially below 5 °C. requires very short or no activation delays. This shows that a fixed activation delay cannot ensure good performance throughout the year and that seasonally adjustable delays are preferable.

For DHW, activation delay has a limited effect, because DHW demand is stable over the year and short-term drops of the storage temperature below 50 °C in principle do not raise sanitary concerns. However, this conclusion is not valid for systems in which SH and DHW are produced alternately by the same set of HPs (as in the P&D project). In such configurations, long DHW cycles (above 60 min) can induce discomfort on the SH side. Consequently, using one fixed activation delay for both services is not appropriate; instead, the delay should be adapted to the service (SH vs. DHW) and season.

Sensitivity analysis shows that increasing storage volume can partially mitigate these issues, whereas oversizing the HP alone does not resolve the trade-off between cycle duration and comfort.

While the results provide preliminary insight into HP cascading in large, non-retrofitted MFBs, future work should address: i) Effective discomfort in terms of SH should include simulation of the thermal response of building to the related distribution temperature; ii) Similarly, the sanitary consequences of temporary DHW drops below 50°C should be analyzed and discussed with a bacteriologist; (iii) the influence of cycling-related losses on COP, since the simulations do not model cycling penalties.

Finally, the deployment of large ASHPs in existing MFBs is currently constrained by the limited availability of >50 kW models, together with the lack of cascade-control strategies adapted to these installations. To facilitate broader market adoption, future large-capacity ASHPs should integrate cascade management directly, without external MCR systems. While some products already respond partially to these requirements, further developments are necessary to ensure robust system integration, satisfaction of SH/DHW temperature targets and reduced HP wear.

Acknowledgements

This research was co-funded by Services Industriels de Genève, the Swiss Federal Office of Energy (within the “Cascade” project), as well as Innosuisse–Swiss Innovation Agency (within the “Renowave” project).

References

[1] Ilic SM, Bullard CW, Hrnjak PS. Effect of Shorter Compressor On/Off Cycle Times on A/C System Performance. Air Cond Refrig Cent CR-43 2001.

[2] Li H, Xu Z, Yang Q, Shao S, Wang Z, Xu C, et al. Enhancing partial load efficiency of air source heat pump heating systems: A comprehensive review of magnitude, influencing factors, and improvement strategies. J Build Eng 2024;97:110715. https://doi.org/10.1016/j.jobe.2024.110715.

[3] Piechurski K, Szulgowska-Zgrzywa M, Danielewicz J. The impact of the work under partial load on the energy efficiency of an air-to-water heat pump. E3S Web Conf 2017;17:00072. https://doi.org/10.1051/e3sconf/20171700072.

[4] Bagarella G, Lazzarin R, Noro M. Sizing strategy of on–off and modulating heat pump systems based on annual energy analysis. Int J Refrig 2016;65:183–93. https://doi.org/10.1016/j.ijrefrig.2016.02.015.

[5] Fraga C, Hollmuller P, Schneider S, Lachal B. Heat pump systems for multifamily buildings: Potential and constraints of several heat sources for diverse building demands. Appl Energy 2018;225:1033–53. https://doi.org/10.1016/j.apenergy.2018.05.004.

[6] Montero O, Brischoux P, Callegari S, Fraga C, Rüetschi M, Vionnet E, et al. Large Air-to-Water Heat Pumps for Fuel-Boiler Substitution in Non-Retrofitted Multi-Family Buildings—Energy Performance, CO2 Savings, and Lessons Learned in Actual Conditions of Use. Energies 2022;15:5033. https://doi.org/10.3390/en15145033.

[7] Montero O, Brischoux P, Fraga C, Rüetschi M, Calame N, Rognon F, et al. Monovalent and Hybrid Air-source Heat Pump Concepts for Existing Multifamily Buildings. 14th IEA Heat Pump Conf 2023 2022.